

Aus dem Max von Pettenkofer-Institut für Hygiene und Medizinische
Mikrobiologie der Ludwig-Maximilians-Universität München

Lehrstuhl: Bakteriologie

Kommissarische Leitung:

Prof. Dr. Rainer Haas

**"Untersuchungen des horizontalen
Gentransfers in *Enterobacteriaceae* durch
Next-Generation Sequencing"**

Dissertation

zum Erwerb des Doktorgrades der Naturwissenschaften

an der Medizinischen Fakultät

der Ludwig-Maximilians-Universität München

vorgelegt von

Maxim Messerer

aus Steingaden

Jahr

2016

**Gedruckt mit Genehmigung der Medizinischen Fakultät
der Ludwig-Maximilians-Universität München**

Betreuer: Prof. Dr. Wolfgang Fischer

Zweitgutachter: Prof. Dr. Kai Papenfort

Dekan: Prof. Dr. med. dent. Reinhard Hickel

Tag der mündlichen Prüfung: 21.12.2016

...aux gens qui sont importants pour moi!

Inhaltsverzeichnis

Inhaltsverzeichnis	- 5 -
Abkürzungsverzeichnis	- 8 -
A. Einleitung	- 12 -
1. <i>Enterobacteriaceae</i>	- 12 -
2. <i>Escherichia coli</i>	- 12 -
2.1. ECOR-Sammlung.....	- 13 -
2.2. Die Anfänge der Klassifizierung von <i>E. coli</i> -Isolaten	- 13 -
2.3. Phylogenetische Gruppen.....	- 14 -
2.4. Multilocus-Sequenz-Typisierung (MLST).....	- 14 -
2.5. Triplex-PCR	- 17 -
2.6. Virulenzfaktoren der ExPEC.....	- 18 -
3. Horizontaler Gentransfer (HGT)	- 19 -
3.1. Transduktion.....	- 20 -
3.2. Transformation	- 20 -
3.3. Konjugation	- 21 -
4. Genomische Inseln (GEIs) und Pathogenitätsinseln (PAIs)	- 22 -
4.1. <i>High-Pathogenicity Island</i> (HPI)	- 23 -
4.2. <i>pks</i> -Insel	- 25 -
4.3. <i>serU</i> -Insel	- 26 -
4.4. PAI III	- 27 -
5. <i>Next-Generation Sequencing</i>	- 28 -
6. Ziele dieser Arbeit	- 31 -
B. Material und Methoden	- 32 -
1. Geräte	- 32 -
2. Chemikalien, Enzyme, Puffer und kommerzielle Labor-Kits	- 33 -
3. Nährmedien	- 34 -
3.1. Flüssigmedien.....	- 34 -
3.2. Festmedien	- 35 -
4. Antibiotika.....	- 37 -
5. Bakterienstämme	- 38 -

6. Plasmide	- 42 -
7. Oligonukleotide	- 43 -
8. Software	- 44 -
9. Mikrobiologische Methoden	- 45 -
9.1. Allgemeine Kultivierung von Bakterien	- 45 -
9.2. Transformation und Selektion	- 45 -
9.3. Konjugation und Selektion	- 46 -
10. Molekularbiologische und genetische Methoden	- 50 -
10.1. Isolierung von DNA für die PCR	- 50 -
10.2. Isolierung von Plasmiden	- 51 -
10.3. Polymerase-Kettenreaktion (PCR)	- 51 -
10.4. Trennung, Aufreinigung und Konzentrationsbestimmung von Nukleinsäuren ..	- 52 -
10.5. Verdau von Nukleinsäuren mit Restriktionsendonukleasen	- 53 -
10.6. Insertion von Antibiotika-Resistenzkassetten	- 53 -
10.7. Sequenzierung von kleinen DNA Abschnitten	- 54 -
11. <i>Next-Generation Sequencing</i> (NGS)	- 54 -
11.1. Auswahl und Anzucht der Bakterien	- 54 -
11.2. Isolierung und Qualitätsüberprüfung der genomischen DNA (gDNA)	- 55 -
11.3. Überprüfung auf Verdaubarkeit der genomischen DNA	- 55 -
11.4. Sequenzierung	- 55 -
11.5. Verwendung der Software CLC Genomics Workbench	- 56 -
C. Ergebnisse	- 59 -
1. <i>Next-Generation Sequencing</i>	- 60 -
2. Ergebnisse der durchgeführten PCRs und Einteilung in PAI-Gruppen	- 60 -
3. Phylogenetische Auswertungen der <i>Housekeeping</i> -Gene und der immobilen PAIs ..	- 64 -
3.1. <i>Housekeeping</i> -Gene	- 65 -
3.2. <i>High-Pathogenicity Island</i> (HPI)	- 72 -
3.3. <i>pks</i> -Insel	- 76 -
3.4. <i>serU</i> -Insel	- 79 -
3.5. Inter-PAI-Regionen	- 82 -
3.6. PAI III	- 86 -
3.7. Untersuchung des Stammes ECOR65	- 98 -
3.8. Untersuchung weiterer ECOR-Stämme	- 101 -
4. Transfer der immobilen PAIs	- 101 -

4.1. Erstellung der Donoren und Konjugation	- 101 -
4.2. <i>in vitro</i> -Auswertung der Transkonjuganten und Konjugationseffizienz.....	- 103 -
4.3. <i>in silico</i> -Auswertung der Transkonjuganten	- 105 -
5. Phylogenetische Auswertung der mobilen PAIs	- 109 -
5.1. Konservierte Kern-Region	- 109 -
5.2. Region X	- 113 -
5.3. Region III: <i>pks</i> -Insel.....	- 115 -
6. Transfer der mobilen PAI.....	- 118 -
6.1. Konstruktion der Donoren und Konjugation.....	- 118 -
6.2. Auswertung der Transkonjuganten	- 119 -
D. Diskussion	- 123 -
1. <i>Next-Generation Sequencing</i>	- 124 -
2. Vergleich der PCR Ergebnisse	- 125 -
3. <i>Housekeeping</i> -Gene und phylogenetische Gruppen.....	- 126 -
4. Pathogenitätsinseln.....	- 130 -
4.1. Phylogenie und Transfer der immobilen Pathogenitätsinseln.....	- 130 -
4.2. Phylogenie und Transfer der mobilen Pathogenitätsinseln	- 140 -
5. Schlussfolgerung und Übertragungsmodell	- 145 -
6. Ausblick	- 146 -
Zusammenfassung	- 147 -
Summary	- 149 -
Anhang	- 151 -
Abbildungsverzeichnis	- 151 -
Tabellenverzeichnis.....	- 153 -
Literaturverzeichnis.....	- 155 -
Danksagung	- 165 -
Eidesstattliche Versicherung	- 167 -

Abkürzungsverzeichnis

A	Adenin
Abb.	Abbildung
ad	Auffüllen auf
Ap	Ampicillin
Ap ^R	Ampicillin-Resistenz
<i>asn</i>	Asparagin
BLAST	engl.: <i>Basic Local Alignment Search Tool</i>
bp	Basenpaare
C	Cytosin
CAS	Chrom Azurol S
Cm	Chloramphenicol
Cm ^R	Chloramphenicol-Resistenz
DNA	Desoxyribonukleinsäure (engl.: <i>deoxyribonucleic acid</i>)
DR	Wiederholungssequenzen (engl.: <i>direct repeats</i>)
ECOR	<i>Escherichia coli</i> Referenz
engl.	englisch
<i>et al.</i>	und andere (lat.: <i>et alii, et aliae</i> oder <i>et alia</i>)
ExPEC	extraintestinal pathogene <i>Escherichia coli</i>
F	Fertilität
for	Vorwärtsprimer (engl.: <i>forward primer</i>)
G	Guanin
gDNA	genomische DNA
GEI	genomische Insel
Hfr	hohe Rekombinationsfrequenz (engl.: <i>high frequency of recombination</i>)
HGT	horizontaler Gentransfer

HPI	engl.: <i>High-Pathogenicity Island</i>
H ₂ O _{dest}	destilliertes Wasser
ICE	integratives und konjugatives Element (engl.: <i>integrative and conjugative element</i>)
IPEC	intestinal pathogene <i>Escherichia coli</i>
IS	Insertions-Sequenz
Kat. #	Katalognummer
kb	Kilobasenpaare
Kn	Kanamycin
Kn ^R	Kanamycin-Resistenz
l	Liter
lat.	lateinisch
LB	engl.: <i>lysogeny broth</i>
log	Logarithmus zur Basis 10
Mb	Megabasen
min	Minute
MLEE	Multilocus-Enzym-Elektrophorese
MLST	Multilocus-Sequenz-Typisierung
NCBI	engl.: <i>National Center for Biotechnology Information</i>
NF-κB	engl.: <i>nuclear factor kappa-light-chain-enhancer of activated B-cells</i>
NGS	engl.: <i>Next-Generation Sequencing</i>
nt	Nukleotid
OD	optische Dichte
ori	Replikationsursprung (engl.: <i>origin of replication</i>)
p	Plasmid
PAI	Pathogenitätsinsel

PBS	phosphatgepufferte Salzlösung (engl.: <i>phosphate-buffered saline</i>)
PCR	Polymerase-Kettenreaktion (engl.: <i>polymerase chain reaction</i>)
phyl.	phylogenetisch
RegX	Region X
rev	Rückwärtsprimer (engl.: <i>reverse primer</i>)
RFLP	Restriktionsfragmentlängenpolymorphismus
rpm	Umdrehungen pro Minute (engl.: <i>rounds per minute</i>)
RT	Raumtemperatur
<i>ser</i>	Serin
SNP	Einzelnukleotid-Polymorphismus (engl.: <i>single nucleotide polymorphism</i>)
T	Thymin
Tab.	Tabelle
TAE	Tris-Acetat-EDTA
TcpC	TIR-Domänen-enthaltendes-Protein <i>coli</i> (engl.: <i>TIR domain containing protein coli</i>)
<i>thr</i>	Threonin
TIR	TOLL/IL-1 Rezeptor
TLR	Toll-like Rezeptor
tRNA	Transfer-Ribonukleinsäure (engl.: <i>transfer ribonucleic acid</i>)
u	Enzymeinheit (engl.: <i>unit</i>)
üN	über Nacht
UPEC	uropathogene <i>Escherichia coli</i>
UV	Ultraviolett
V	Volt

VNTR	Tandem-Region (engl.: <i>variable numbers of tandem repeats</i>)
v/v	Volumen zu Volumen
WT	Wildtyp
w/v	Gewicht zu Volumen (engl.: <i>weight/volume</i>)
Z!	Zentrifugationsschritt

A. Einleitung

1. Enterobacteriaceae

Die Familie *Enterobacteriaceae* innerhalb der Domäne Bakterien umfasst eine Vielzahl unterschiedlicher Arten. Der Name dieser Familie leitet sich von dem altgriechischen Wort *enteron* ab, was übersetzt Darm bedeutet. Dieser Name beruht darauf, dass viele Vertreter der *Enterobacteriaceae* typische Darmbewohner sind. Sie sind stäbchenförmig, Gram-negativ und besitzen meist Flagellen zur Fortbewegung. Bekannte Vertreter dieser Familie sind z.B. *Klebsiella pneumoniae*, *Yersinia enterocolitica*, *Citrobacter koseri* und *Escherichia coli*, die sowohl als Kommensale als auch als Pathogene zu finden sind (Williams et al., 2010). Aufgrund der Besiedelung des menschlichen Darms durch pathogene Vertreter wurden *Enterobacteriaceae* bereits intensiv molekulargenetisch und infektionsimmunologisch untersucht. Auch als Erreger Krankenhaus-assoziiierter Infektionen rückten Vertreter dieser Familie vermehrt in den Fokus des Interesses (Nakai et al., 2016).

2. Escherichia coli

Das Bakterium *Escherichia coli* wurde nach seinem Entdecker Theodor Escherich (1857 - 1911) benannt, der dieser Art den Namen "*Bacterium coli commune*" gab (Escherich, 1989). Es ist eine der am besten beschriebenen Bakterienarten weltweit und der Modellorganismus in der Mikrobiologie und der Genetik. Im Rahmen des *E. coli*-Genomprojekts der Universität von Wisconsin (USA) wurde im Jahre 1997 das gesamte Genom des Stammes *E. coli* MG1655 erstmals komplett sequenziert. Dies schaffte damit eine neue Grundlage der molekularen Infektionsforschung (Blattner et al., 1997). Das bakterielle Genom der Art *E. coli* besitzt eine variable Größe von 4,5 bis 5,5 Megabasen (Mb) (Ochman und Bergthorsson, 1998). Diese Unterschiede im Chromosom kommen dadurch zustande, dass neben lebenswichtigen Genen (Kerngenom) auch weitere akzessorische Genbereiche (flexibles Genom) vorhanden sind, die z.B. für Antibiotikaresistenzen und Virulenzfaktoren codieren. Neben der chromosomalen DNA besitzen viele *E. coli*-Stämme noch extrachromosomale DNA (Plasmide) (Hakenbeck et al., 2001; Hacker et al., 2004; Dobrindt et al., 2010).

Die Art *E. coli* ist ein natürlicher Bestandteil der Darmflora, allerdings können einige Isolate eine Vielzahl an intestinalen und extraintestinalen Krankheiten auslösen (Orskov und Orskov, 1992). Während intestinal pathogene *E. coli* (IPEC) vor allem für Durchfallerkrankungen verantwortlich sind, verursachen extraintestinal pathogene *E. coli* (ExPEC) Infektionen, die von asymptomatischen Harnwegsinfektionen bis zu lebensbedrohlichen Krankheiten wie Pyelonephritis (Nierenbeckenentzündung), Sepsis (Blutvergiftung) und Meningitis (Hirnhautentzündung) bei Neugeborenen reichen (Picard et al., 1999; Johnson und Stell, 2000; Johnson und Russo, 2002).

2.1. ECOR-Sammlung

Die *Escherichia coli* Referenz (ECOR)-Sammlung besteht aus 72 *E. coli*-Stämmen und wurde Anfang der 1980er von Howard Ochman und Robert Selander zusammengestellt. Diese 72 Bakterienstämme wurden aus einer Kollektion von 2.600 untersuchten Isolaten ausgewählt, um die genetische Diversität der Art *E. coli* zu repräsentieren. Hierfür sollten die Stämme verschiedene Kriterien erfüllen. Es wurden Isolate gewählt, die alle verschiedenen Gruppen, die durch Multilocus-Enzym-Elektrophorese (MLEE) definiert wurden, repräsentieren und die bereits in Veröffentlichungen verwendet wurden. Zudem war die Sammlung von verschiedenen Wirten und an verschiedenen globalen Standorten isoliert worden und enthielt sowohl Kommensale als auch Pathogene (Ochman und Selander, 1984). Die ECOR-Sammlung wurde in den letzten Dekaden von vielen Arbeitsgruppen als Referenzsammlung verwendet. Dabei entwickelten sich leichte Unterschiede in den verschiedenen Sammlungen, was durch die unterschiedliche Verteilung der Virulenzfaktoren innerhalb der eigentlich identischen Isolate gezeigt wurde (Johnson et al., 2001a). Es wurden auch Verunreinigungen und Verwechslungen der einzelnen Bakterienstämme in den erwerblichen Sammlungen festgestellt (Schubert, mündliche Mitteilung).

2.2. Die Anfänge der Klassifizierung von *E. coli*-Isolaten

Innerhalb der letzten Jahrzehnte wurden verschiedene Typisierungsansätze zur Klassifizierung von *E. coli*-Isolaten entwickelt, die auch heutzutage noch angewandt werden (Tenailon et al., 2010). Die Serotypisierung für *E. coli* wurde im Jahr 1947 von Fritz Kauffmann vorgestellt. Die Serotypen werden anhand der O- (O-Polysaccharid), K- (Kapsel) und H- (Flagellen) Oberflächen-Antigene definiert (Kauffmann, 1947). Im Jahr 1973

veröffentlichte Roger Milkman eine Studie, in der er die elektrophoretische Mobilität von Enzymen verschiedener *E. coli*-Isolate beschrieb (Milkman, 1973). Aufgrund dieser Beobachtungen wurde die Multilocus-Enzym-Elektrophorese (MLEE) entwickelt, die sich als Standard-Technik für Untersuchungen der bakteriellen Diversität etablierte (Chaudhuri und Henderson, 2012). Die ribosomale Typisierung (*Ribotyping*) wurde im Jahr 1986 beschrieben (Grimont und Grimont, 1986). Diese Technik beruht auf der Methode des Restriktionsfragmentlängenpolymorphismus (RFLP), bei der DNA-Fragmente mit Hilfe von Restriktionsenzymen geschnitten und anschließend durch eine Gelelektrophorese anhand ihrer Länge angeordnet werden. Durch einen Vergleich der entstandenen Fragment-Muster können Bakterien so miteinander verglichen werden (Bouchet et al., 2008). Mit den Methoden MLEE und *Ribotyping* wurde es möglich, die untersuchten *E. coli*-Isolate in phylogenetische Gruppen einzuteilen (Chaudhuri und Henderson, 2012; Bouchet et al., 2008).

2.3. Phylogenetische Gruppen

Phylogenetische Analysen zeigten, dass die Struktur der *E. coli*-Population neben Rekombinations-Ereignissen größtenteils auf Klonalität beruht (Tenaillon et al., 2010). Aufgrund dieser stammesgeschichtlichen Entwicklung können *E. coli*-Isolate in vier phylogenetische Hauptgruppen (A, B1, B2 und D) eingeteilt werden (Herzer et al., 1990; Clermont et al., 2000). ExPEC sind in der phylogenetischen Gruppe B2, aber auch teilweise in der Gruppe D zu finden, während IPEC hauptsächlich zu den Gruppen A und B1 gehören (Jaureguy et al., 2008; Toval et al., 2014). Die Gruppen A und B1 sind sogenannte Schwester-Gruppen, also untereinander enger verwandt als zu den anderen Gruppen (Clermont et al., 2000). Es wurde von einer phylogenetischen Untergruppe E berichtet, die aber meist vernachlässigt wird, weil sie sich inkonsistent in den Verwandtschaftsanalysen einreicht (Wirth et al., 2006). Zusätzlich wurde auch eine Untergruppe F beschrieben, die allerdings vielfach der phylogenetischen Hauptgruppe D zugerechnet wird (Jaureguy et al., 2008).

2.4. Multilocus-Sequenz-Typisierung (MLST)

Die Multilocus-Sequenz-Typisierung (MLST) wurde in den späten 1990er Jahren eingeführt (Enright und Spratt, 1999). Für die Einteilung von Bakterienstämmen in verschiedene phylogenetische Gruppen wird diese Technik heutzutage am häufigsten genutzt. Bei der MLST werden die DNA-Sequenzen von verschiedenen Loci aus dem bakteriellen Genom von

Isolaten derselben Art miteinander verglichen. Die Loci sind Fragmente sogenannter *Housekeeping*-Gene, die hoch konserviert sind, aber trotzdem eine gewisse Variabilität zwischen unterschiedlichen *E. coli*-Isolaten aufweisen. Diese Fragmente weisen meist eine Größe von ca. 500 bp auf (Maiden et al., 1998; Sorokin et al., 2006; Sahl et al., 2012). Traditionell müssen die Gen-Fragmente eines Isolats erst per PCR amplifiziert und nachfolgend sequenziert werden. Bei einer hohen Anzahl an untersuchten Stämmen kann dies recht zeit- und kostenintensiv sein (Larsen et al., 2012). Die erhaltenen Daten werden anschließend *in silico* zu längeren Sequenzen aneinandergereiht und mit denen der anderen Isolate verglichen. Mit der relativ jungen Methode der Hochdurchsatz-Sequenzierung, deren Kosten momentan jährlich sinken, kann der aufwendige Schritt der PCR-Amplifizierung umgangen werden (Larsen et al., 2012). Zur Darstellung der Verwandtschaft werden meist phylogenetische Stammbäume unter Verwendung von bestimmten Algorithmen erstellt (Lyubetsky et al., 2016). Es existieren verschiedene MLST-Schemata, die die Fragmente von teils unterschiedlichen *Housekeeping*-Genen analysieren. Häufig verwendete Schemata sind das Achtman- und das Pasteur-Schema (Wirth et al., 2006; Jaureguy et al., 2008).

Das Achtman-Schema wird für verschiedene Bakterienarten angewandt und wurde von Mark Achtman (Universität Warwick, England) entworfen. Für die phylogenetische Klassifizierung von *E. coli* werden die Sequenzen von Fragmenten der sieben Gene *adk* (Adenylat-Kinase), *icd* (Isocitrat/Isopropylmalat-Dehydrogenase), *fumC* (Fumarat-Hydratase), *recA* (ATP/GTP-Bindemotif), *mdh* (Malat-Dehydrogenase), *gyrB* (DNA-Gyrase) und *purA* (Adenylosuccinat-Dehydrogenase) verwendet. Die Lokalisation dieser Gene im *E. coli*-Genom ist schematisch in Abbildung 1 zu sehen. Die hier verwendeten *Housekeeping*-Gene sind über das gesamte bakterielle Genom verteilt (Wirth et al., 2006).

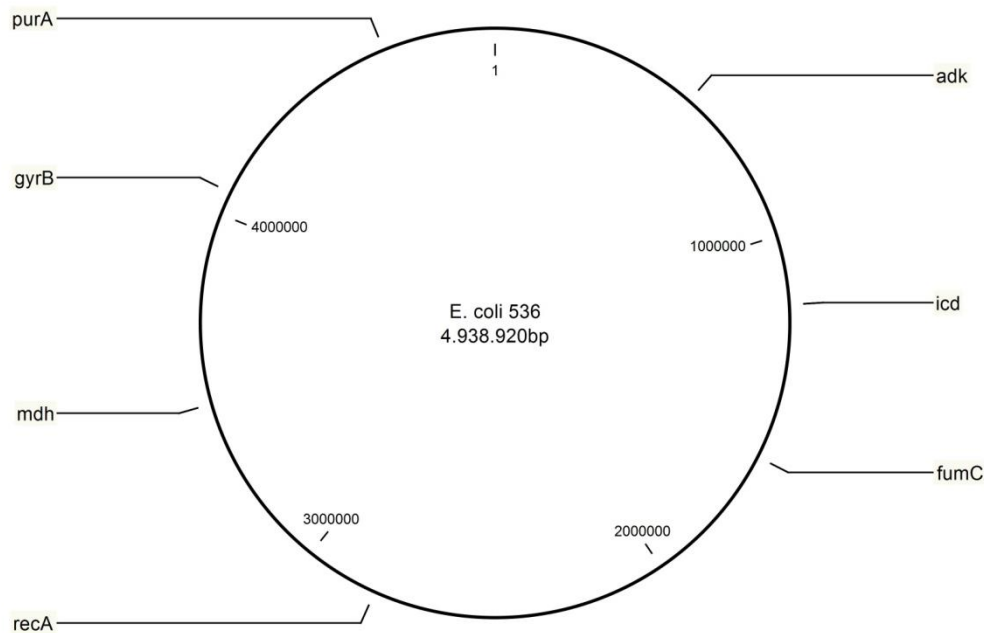


Abbildung 1: Schematische Darstellung des Genoms des Stammes *E. coli* 536. Gezeigt werden die sieben Gene *adk*, *icd*, *fumC*, *recA*, *mdh*, *gyrB* und *purA*, deren Fragmente für das Achtman MLST-Schema verwendet werden.

Das vom Pasteur-Institut (Paris, Frankreich) entwickelte Pasteur-Schema wird für verschiedene Bakterienarten verwendet. Für die Einteilung von *E. coli* in die phylogenetischen Gruppen werden Fragmente der acht Gene *polB* (Polymerase PolII), *dinB* (DNA-Polymerase), *putP* (Prolin-Permease), *icdA* (Isocitrat-Dehydrogenase), *trpA* (Tryptophan-Synthase Untereinheit A), *trpB* (Tryptophan-Synthase Untereinheit B), *uidA* (Beta-Glucuronidase) und *pabB* (p-Aminobenzoat-Synthase) verglichen (Jaureguy et al., 2008). Die Lokalisation der Gene im *E. coli*-Genom ist in Abbildung 2 zu sehen. Es existiert auch ein modifiziertes Schema, das nur die Fragmente von sechs der acht Gene beinhaltet (*trpA*, *trpB*, *pabB*, *putP*, *icdA* und *polB*) (Escobar-Paramo et al., 2004). Es wird dem Pasteur-Schema zugesprochen, dass die verwendeten Gene unter geringem Einfluss von horizontalem Gentransfer stehen und die Einteilung in die phylogenetischen Gruppen somit der Klonalität der Art entspricht (Lecointre et al., 1998; Denamur et al., 2000).

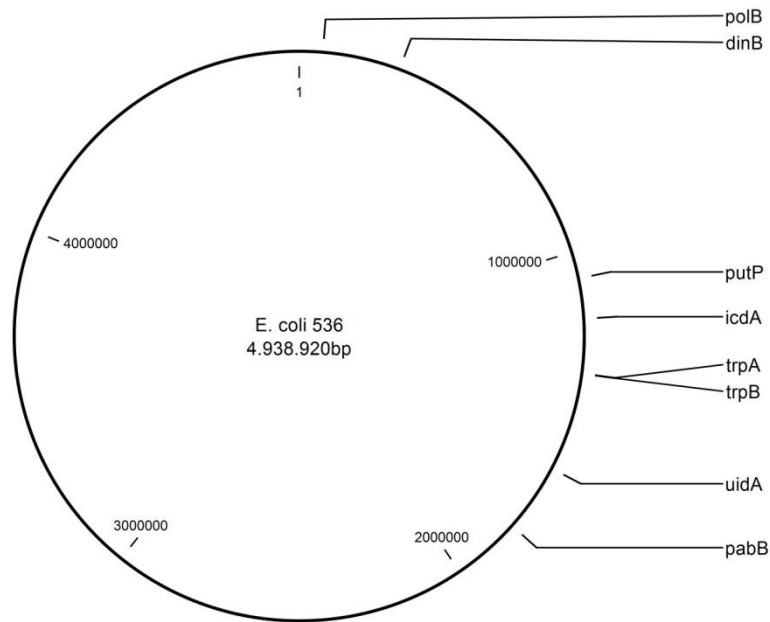


Abbildung 2: Schematische Darstellung des Chromosoms des *E. coli*-Stammes 536. Dargestellt sind die acht Gene *polB*, *dinB*, *putP*, *icdA*, *trpA*, *trpB*, *uidA* und *pabB*, deren Fragmente für das Pasteur MLST-Schema verwendet werden.

2.5. Triplex-PCR

Auf der Suche nach einer vereinfachten Methode zur Einteilung der *E. coli*-Stämme in die vier phylogenetischen Hauptgruppen haben Clermont *et al.* ein Verfahren vorgestellt, dessen Grundlage auf einer Polymerase-Kettenreaktion (PCR) basiert. In diesem Fall handelt es sich um eine Triplex-PCR mit drei verschiedenen Primer-Paaren, die jeweils ein Amplifikat erzeugen können. Aus dem Muster der An- bzw. Abwesenheit der PCR-Amplifikate ist, wie in Abbildung 3 dargestellt, eine einfache Einteilung der Isolate in die vier phylogenetischen Hauptgruppen möglich. Da die drei Fragmente unterschiedliche Größen besitzen, kann durch Betrachtung des nachfolgenden Gelbilds die Gruppe bestimmt werden. Wenn z.B. das Ergebnis der Triplex-PCR eines Isolats eine Bande der Größe 279 Basenpaare (bp) (*chuA*) und eine der Größe 211 bp (*yjaA*) aufweist, handelt es sich um einen Vertreter der phylogenetischen Gruppe B2. Ein weiteres Beispiel wäre das Vorhandensein von nur einer Bande der Größe 152 bp (*TspE4.C2*), was einen Stamm der Gruppe B1 identifizieren würde. Somit ist eine rasche Einteilung in eine der vier Hauptgruppen möglich (Clermont *et al.*, 2000). Allerdings kann mit dieser Methode nicht der Grad der Verwandtschaft bestimmt werden.

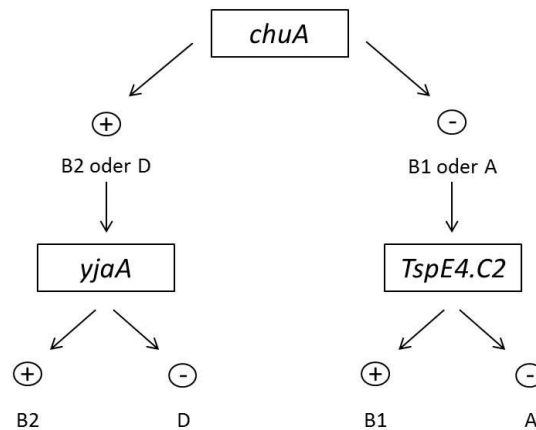


Abbildung 3: Schematische Darstellung der Auswertung der Triplex-PCR. Es wird das Vorhandensein von drei verschiedenen Gen-Fragmenten überprüft. Somit kann ein *E. coli*-Isolat sehr schnell in eine der vier Hauptgruppen A, B1, B2 und D eingeordnet werden.

2.6. Virulenzfaktoren der ExPEC

Die Pathogruppe ExPEC ist für die meisten Infektionen außerhalb des Darms verantwortlich, wofür unter anderem deren Virulenzfaktoren eine wichtige Rolle spielen (Picard et al., 1999; Johnson und Stell, 2000; Johnson und Russo, 2002). Die meisten Virulenzfaktoren sind innerhalb der phylogenetischen Gruppen B2, D und teilweise auch B1 zu finden, während diese innerhalb der Gruppe A weniger präsent sind (Herzer et al., 1990; Clermont et al., 2000; Johnson et al., 2001a). Die Verteilung der Virulenzfaktoren ist auch anhand der Genomgröße der *E. coli*-Stämme ersichtlich. Während die durchschnittliche Genomgröße der phylogenetischen Gruppen B1, B2 und D $5,07 \pm 0,09$ Mb beträgt, ist diese bei der Gruppe A $4,74 \pm 0,06$ Mb (Bergthorsson und Ochman, 1998). Es wurde gezeigt, dass sich die meisten ExPEC-Stämme aus *E. coli*-Stämmen durch eine etappenweise Ansammlung von Virulenzfaktoren durch horizontalen Gentransfer (HGT) entwickelt haben (Dobrindt und Hacker, 2001; Bidet et al., 2005).

Obwohl ExPECs ca. 20% der fäkalen *E. coli*-Stämme repräsentieren, verursachen sie keine Krankheiten im Magen-Darm-Trakt (Siitonen, 1992; Johnson et al., 1998; Murray et al., 2004; Nowrouzian et al., 2005). Ihre Virulenzfaktoren ermöglichen den ExPECs allerdings, extraintestinal die oberflächliche Schleimhaut des Wirts zu besiedeln, wichtige Nährstoffe wie Eisen zu erhalten und das Immunsystem zu untergraben oder diesem auszuweichen (Johnson, 1991; Bower et al., 2005; Cirl et al., 2008; Meyer et al., 2015).

Eisen ist ein wichtiger Nährstoff für *E. coli* und die meisten anderen Mikroorganismen. Obwohl Eisen ein sehr häufig vorkommendes Element ist, kann es nur in Form von

Eisenionen (Fe^{2+} bzw. Fe^{3+}) von den Bakterien genutzt werden. Im Wirt ist das verfügbare Eisen zudem meist intrazellulär vorhanden und fest an Proteine gebunden, wie z.B. im Hämoglobin. Um an das benötigte Eisen zu gelangen, besitzen ExPECs verschiedene Eisenaufnahmesysteme, wie z.B. Siderophore (Neilands, 1976; Crosa, 1989; Raymond et al., 2003). Bekannte Vertreter sind Aerobaktin, Enterobaktin, Salmochelin und Yersiniabaktin. Siderophore sind eisenbindende Moleküle, die sekretiert werden, um sowohl freies als auch durch Wirtsproteine gebundenes Eisen aufzunehmen. Das Bakterium kann anschließend über spezielle Rezeptoren wie IroN (Salmochelin) und FyuA (Yersiniabaktin) die Siderophore mit dem gebundenen Eisen aufnehmen (Hantke et al., 2003; Hancock et al., 2008; Wiles et al., 2008).

Siderophor-Systeme und viele Virulenzfaktoren sind meist auf Pathogenitätsinseln (PAIs) lokalisiert und können über HGT weitergegeben werden (Groisman und Ochman, 1996; Dobrindt und Hacker, 2001).

3. Horizontaler Gentransfer (HGT)

Die Vermehrung von Bakterien geschieht durch Zellteilung. Die Weitergabe von genetischen Informationen auf die Tochterzellen wird deshalb auch als vertikaler Gentransfer bezeichnet. Horizontaler Gentransfer (HGT) beschreibt dagegen die Übertragung von Genen, die nicht mit einer Zellteilung einhergeht. Dies erfolgt normalerweise durch Transformation, Transduktion und Konjugation und kann auch über die Gattungsgrenze hinaus geschehen. Hierfür sind meist Vehikel notwendig, wie z.B. Phagen, Plasmide oder Insertionssequenz (IS)-Elemente (Avery et al., 1979; Lederberg und Tatum, 1953; Zinder und Lederberg, 1952). Ein schematischer Vergleich zwischen vertikaalem und horizontalem Gentransfer ist in Abbildung 4 zu sehen. Mit der Möglichkeit der Genomsequenzierung kann erworbene DNA in einem Organismus identifiziert werden. Das Bakterium *E. coli* hat schätzungsweise zwischen 8 und 18% seines Genoms durch HGT erhalten (Ochman et al., 2000; Lawrence, 1999).

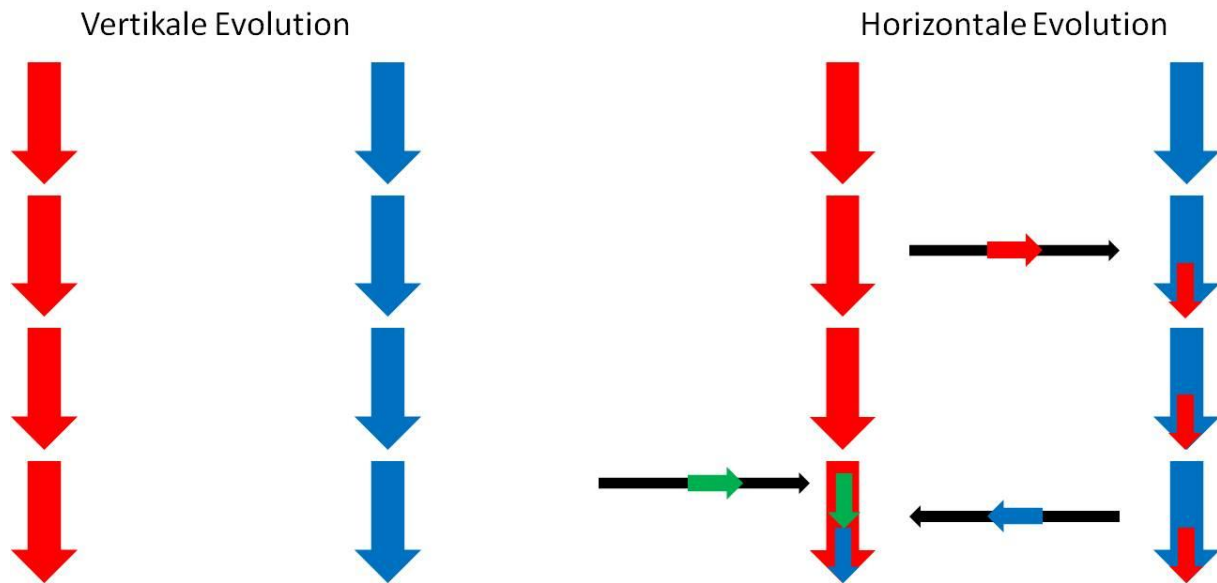


Abbildung 4: Darstellung des vertikalen und horizontalen Gentransfers bei Bakterien. Bei der vertikalen Evolution werden die genetischen Informationen durch Zellteilung an die Tochterzellen weitergegeben. Bei der horizontalen Evolution kommen zusätzlich Gene von anderen Organismen hinzu, die sich im Genom festigen können und über den vertikalen Weg weiter gegeben werden.

3.1. Transduktion

Transduktion ist die Weitergabe von genetischem Material durch Phagen. Es gibt zwei verschiedene Formen, die unspezifische und die spezifische Transduktion. In beiden Fällen wird bei der Infektion des Bakteriums ein Teil von dessen DNA in den Phagen eingebaut und kann danach auf ein weiteres Bakterium übertragen werden. Die spezifische Transduktion wurde bereits im Jahr 1947 durch Joshua Lederberg und Edward Lawrie Tatum und im Jahr 1952 die unspezifische ebenfalls durch Lederberg und seinem Schüler Norton Zinder nachgewiesen (Zinder und Lederberg, 1952; Lederberg und Tatum, 1953).

3.2. Transformation

Transformation beschreibt die Aufnahme freier DNA durch Bakterien direkt aus der Umgebung. Es war der erste Gen-Transfer-Mechanismus, der beschrieben wurde. Im Jahr 1928 zeigte Frederick Griffith anhand von Pneumokokken-Stämmen, dass die Befähigung der Kapselbildung übertragen werden kann (Griffith, 1928). Einige Jahre später bewies O. Avery, dass dies durch den Transfer von DNA zustande kam (Avery et al., 1979). In der Mikrobiologie wird die Transformation dazu genutzt, DNA-Fragmente oder Plasmide in ein Bakterium zu transferieren.

3.3. Konjugation

Bei der Konjugation wird DNA von einem bakteriellen Donor auf einen Rezipienten transferiert. Dies geschieht durch den F-Pilus des Donors, auch bekannt unter dem Namen Sex-Pilus, der damit eine Verbindung zu dem Rezipienten aufbaut. F steht hier für Fertilität. Hierfür müssen die beiden Bakterienzellen in engem Kontakt stehen (Frost et al., 1994). Die Konjugation wurde im Jahr 1946 durch Joshua Lederberg und Edward Lawrie Tatum nachgewiesen (Adelberg und Pittard, 1965).

Es können Plasmide oder auch ganze Genomabschnitte übertragen werden. Damit ein Bakterium zur Konjugation befähigt wird, werden Gene für die Mobilisierung und den Transfer benötigt, die meistens auf einem Plasmid codiert sind. Ein bekannter Vertreter ist das F-Plasmid, welches eine Größe von ca. 100 kb besitzt. Dieses Plasmid kann sich selbst replizieren und somit vertikal in einer Zellpopulation weitergegeben werden. Eine Kopie des F-Plasmids kann horizontal auf einen Rezipienten übertragen werden, was diesen zu einem Donor werden lässt. Der gesamte Prozess dauert in etwa nur fünf Minuten, wobei der ursprüngliche Donor eine Kopie des Plasmids behält. Das F-Plasmid und das Chromosom des Donors besitzen verschiedene IS-Elemente. Wenn in beiden dasselbe spezifische IS-Element vorhanden ist, kann das F-Plasmid in das Chromosom des Bakteriums durch homologe Rekombination integrieren. In diesem Fall kann das gesamte Genom eines Donors auf einen Rezipienten übertragen werden. Bakterien mit einem integrierten F-Plasmid werden als *high frequency of recombination* (Hfr; hohe Rekombinationsfrequenz) -Stämme bezeichnet. Das F-Plasmid verlässt gelegentlich das Chromosom eines Hfr-Stammes und ist anschließend wieder im Zytoplasma des Bakteriums zu finden. In wenigen Fällen nimmt das Plasmid auch chromosomale Gene auf und wird dann als F'-Plasmid bezeichnet. Diese spezifischen Gene des Donors können nun zusammen mit dem F'-Plasmid auf einen Rezipienten übertragen werden, welcher anschließend zwei Kopien dieser Gene besitzt, seine ursprünglichen im Chromosom und die neu transferierten auf dem zytoplasmatischen F'-Plasmid. Durch doppeltes *Crossing-Over* an zwei Regionen, die homolog in der Donor- und der Rezipienten-DNA sind, kann die DNA auf dem F'-Plasmid die chromosomale DNA ersetzen. Somit hat der Rezipient neue genetische Informationen erhalten, die nun fest im Genom fixiert sind und wiederum vertikal weitergegeben werden können (Adelberg und Pittard, 1965; Frost et al., 1994; Griffiths et al., 1999). Eine schematische Darstellung des Transfers mit einem F'-Plasmid ist in Abbildung 5 dargestellt.

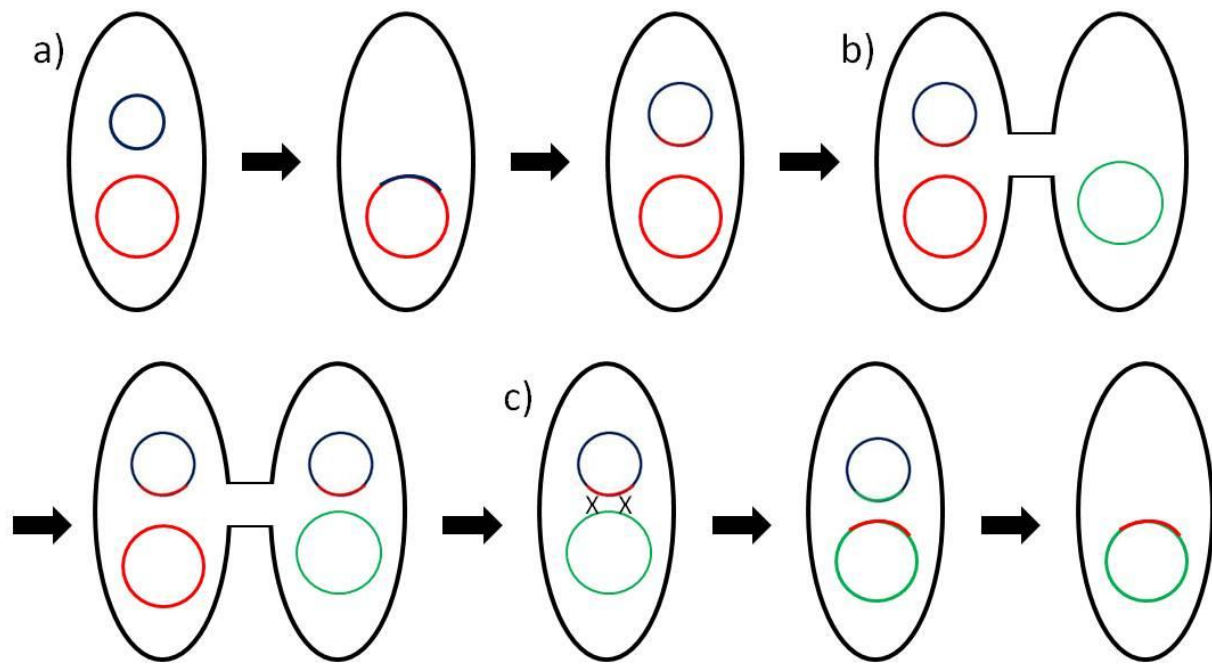


Abbildung 5: Schematische Darstellung einer Übertragung von chromosomaler DNA von einem Donor auf einen Rezipienten durch ein F'-Plasmid. a) Das F-Plasmid (Blau) eines Donors kann sich in dessen Chromosom (Rot) integrieren. Durch inkorrektes Ausschneiden des F-Plasmids kann dieses DNA des Donors aufnehmen, und wird dann als F'-Plasmid bezeichnet. b) Die Übertragung des F'-Plasmids auf einen Rezipienten erfolgt über den F-Pilus, wobei der Donor eine Kopie behält. c) Die Donoren-DNA auf dem F'-Plasmid kann durch doppeltes *Crossing-Over* homologe Bereiche im Rezipienten-Chromosom (Grün) ersetzen. Diese chromosomale DNA des Rezipienten ist dann auf dem F'-Plasmid lokalisiert. Wenn kein Druck besteht, dieses Plasmid zu erhalten, kann es abgebaut werden und geht damit verloren. Somit ist die neu erworbene DNA fest im Chromosom des Rezipienten fixiert und kann vertikal weiter gegeben werden.

4. Genomische Inseln (GEIs) und Pathogenitätsinseln (PAIs)

In vielen bakteriellen Genomen wurden große Regionen gefunden, die durch HGT ausgetauscht wurden. Da sie meist strukturelle Gemeinsamkeiten zeigen, wurden sie als genomische Inseln (GEIs) zusammengefasst. Die GEIs werden ihrem genetischen Inhalt entsprechend in verschiedene Kategorien eingeteilt: symbiotische Inseln, metabolische Inseln, Fitness-Inseln, Resistenz-Inseln und Pathogenitätsinseln (PAIs) (Dobrindt et al., 2004). Die Virulenz eines Bakterienstammes wird meist mit dem Vorhandensein von PAIs in Zusammenhang gebracht (Groisman und Ochman, 1996; Hacker und Kaper, 2000). PAIs besitzen spezielle genetische Eigenschaften (Schmidt und Hensel, 2004):

- i. PAIs tragen ein oder mehrere Virulenzgene.
- ii. PAIs sind in Genomen von pathogenen Bakterien zu finden, aber nicht in denen von artverwandten Apathogenen.

- iii. PAIs besitzen eine durchschnittliche Größe von 10 bis 200 kb.
- iv. PAIs besitzen einen unterschiedlichen G+C-Gehalt im Vergleich zum restlichen Genom des Bakteriums.
- v. PAIs sind meist mit tRNAs assoziiert. Da die tRNAs bei verschiedenen Bakterienarten hoch konserviert sind, können die PAIs auch zwischen den Arten mittels HGT übertragen werden.
- vi. PAIs sind häufig mit mobilen genetischen Elementen verbunden. Oft werden sie von Wiederholungssequenzen (*direct repeats*, DRs) flankiert, welche eine Größe von 16 bis 20 bp besitzen. PAIs tragen häufig kryptische oder noch funktionsfähige Mobilitäts-Gene wie Integrasen oder Transposasen. In PAIs werden oft Insertions-Sequenzen (IS) gefunden. PAIs können auch aus integrierten Plasmiden, konjugativen Transposons oder Bakteriophagen entstanden sein.
- vii. PAIs sind manchmal genetisch instabil. Falls noch funktionsfähige mobile Elemente in einer PAI vorkommen, kann dies auch zum Verlust dieser Insel aus dem bakteriellen Genom führen.
- viii. PAIs zeigen manchmal eine mosaikartige Struktur. Dies könnte daher kommen, dass die erworbene DNA zu verschiedenen Zeitpunkten und von verschiedenen Stämmen in das Genom eines Bakteriums gelangte. Anhand einer Umstrukturierung durch z.B. IS-Elemente könnten sich so manche PAIs entwickelt haben.

Mittlerweile ist eine Vielzahl an PAIs bekannt. In dieser Arbeit wird der Fokus auf vier spezielle PAIs gerichtet, die im Weiteren beschrieben werden: die *High-Pathogenicity Island* (HPI), die *pks*-Insel, die *serU*-Insel und die PAI III.

4.1. High-Pathogenicity Island (HPI)

Ein bisher gut untersuchtes Siderophor-System wird von der *High-Pathogenicity Island* (HPI) synthetisiert. Das Siderophor Yersiniabaktin der HPI wurde zuerst in *Yersinia pestis* entdeckt, wodurch es auch seinen Namen erhielt (Rakin et al., 1994). Die HPI ist in der Familie *Enterobacteriaceae* weit verbreitet und ist vor allem in ExPECs zu finden, die Harnwegsinfektionen, Sepsis und Meningitis in Neugeborenen verursachen (Clermont et al., 2001; Johnson und Stell, 2000; Schubert et al., 1998). Darüber hinaus wird der HPI

zugesprochen, die Virulenz einzelner Stämme stark zu erhöhen (Carniel, 2001; Schubert et al., 2002). Diese Insel ist neben den Arten *E. coli* und *Y. pestis* auch in pathogenen Vertretern der Arten *C. koseri* und *K. pneumoniae* präsent. Allerdings wurde die HPI teilweise auch bei apathogenen Vertretern der Familie *Enterobacteriaceae* nachgewiesen (Lawlor et al., 2007; Bach et al., 2000; Koczura und Kaznowski, 2003; Schubert et al., 2000).

Die Struktur der HPI ist in mehrere Teile gegliedert, welche in Abbildung 6 zu sehen sind. Eine konservierte, G+C-reiche Kern-Region beinhaltet Gene für die Synthese und den Transport des Yersiniabaktins (*irp*) und für den Yersiniabaktin-Rezeptor (*fyuA*), sowie ein Regulator-Gen (*ybtA*). Die linke Seite beinhaltet ein Integrase-Gen, die rechte eine variable A+T-reiche Region, die auch Gene für die Mobilisierung und den Transfer der HPI enthält. Die Insel ist von zwei DRs flankiert und in einer *asn* tRNA inseriert (Schubert et al., 2004b). Das Gen *irp6* ist auch unter dem Namen *ybtQ* bekannt (Benedek und Schubert, 2007). In 99% der HPI-positiven *E. coli*-Stämme ist eine Deletion *downstream* des Gens *fyuA* zu finden, wodurch einerseits die Gene für die Mobilisierung und den Transfer und andererseits das zweite DR fehlen. Dieser HPI-Typ ist somit nicht mehr aktiv transferierbar (Schubert et al., 2004a). Es konnte allerdings gezeigt werden, dass die HPI zusammen mit dem umliegenden chromosomalen *Backbone* mit Hilfe eines F'-Plasmids auf einen weiteren *E. coli*-Stamm übertragen werden kann (Schubert et al., 2009).

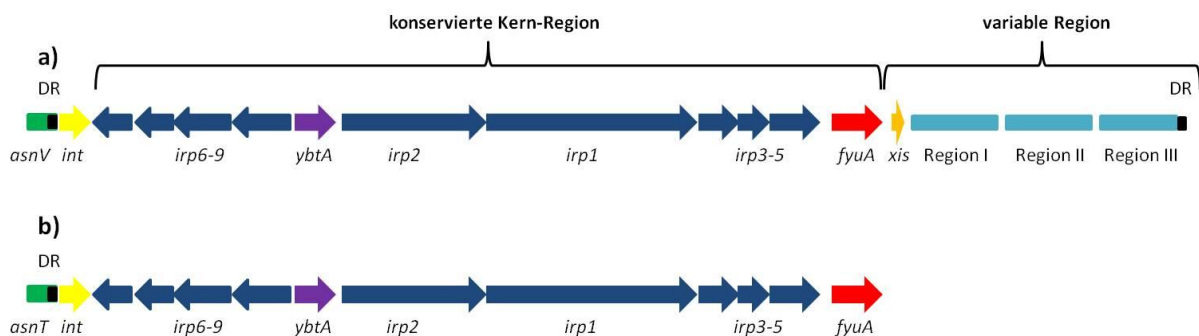


Abbildung 6: Schematische Darstellung der zwei HPI-Typen von a) ECOR31 und b) den meisten HPI-positiven *E. coli*-Stämmen. Beide HPIs sind in einer *asn* tRNA inseriert, aber a) in der *asnV* tRNA und b) in der *asnT* tRNA. Die Gene für die Synthese und den Transport des Yersiniabaktins (*irp*) sind in Blau dargestellt, der Yersiniabaktin-Rezeptor (*fyuA*) in Rot und das Regulator-Gen (*ybtA*) in Lila. In den meisten HPI-positiven *E. coli*-Stämmen fehlt die variable Region.

Die noch aktiv übertragbare Variante ohne Deletion wird als integratives und konjugatives Element (*integrative and conjugative element*, ICE) bezeichnet und wurde in der Art *E. coli*

erstmal bei dem Stamm ECOR31 gefunden. Die Bezeichnung kommt daher, dass ein ICE, ausgehend von einer episomalen Ringstruktur, über Konjugation an ein anderes Bakterium weitergegeben und sich in dessen Genom integrieren kann (Schubert et al., 2004a). Der variable Bereich auf der rechten Seite der ICE-Typ-HPI wird in drei Regionen aufgeteilt. Die Region I beinhaltet die Gene für den Transfer, die Region II die Gene für die Mobilisierung der Insel. Die Region III beinhaltet genetisches Material mit meist unbekannter Funktion (Schubert et al., 2004b). In dieser Region wurden bei ICE-Typ-HPIs einiger Stämme, darunter auch *C. koseri*- und *K. pneumoniae*-Stämme, eine fast vollständig vorhandene PAI gefunden, die *pks*-Insel (Putze et al., 2009).

Es konnte gezeigt werden, dass die ICE-Typ-HPI des Stammes ECOR31 eine Region am Integrase-Ende der HPI besitzt, die hohe Homologien zu dem Virulenz-Plasmid pLVPK von *K. pneumoniae* besitzt. Diese unbekannte Region wurde als Region X betitelt und ebenfalls zusammen mit der HPI in dem *K. pneumoniae*-Stamm B-2495 gefunden, der aus der Blutkultur eines Patienten mit Sepsis stammt. Es wurden in dieser Region auch putative Aufnahmesysteme für divalente Metallionen, wie Eisen, Zink und Mangan gefunden. Zusätzlich wurden Regulator-Gene mit einer Ähnlichkeit zum Eisenaufnahme-Regulator (*Fur*, *Ferric uptake regulator*) entdeckt (Preisach, 2008). Es wird vermutet, dass das Plasmid pLVPK durch verschiedene horizontal erworbene DNA-Fragmente entstand. Auf dem Plasmid wurden neben einigen IS-Elementen auch Eisenaufnahmesysteme wie der *iroBCDN*-Gen-Cluster gefunden und ebenso Gene, die dem Partitionierungssystem von F-Plasmiden ähneln (Chen et al., 2004). Es wurden auch Genabschnitte mit hoher Homologie zu dem Plasmid pLVPK in Verbindung mit der ICE-Typ-HPI in *C. koseri*- und *K. pneumoniae*-Stämmen beschrieben (Putze et al., 2009).

4.2. *pks*-Insel

Bei der Infektion von eukaryotischen Zellen mit gewissen *E. coli*-Stämmen wurde festgestellt, dass bei direktem Kontakt Doppelstrangbrüche in der DNA dieser Zellen auftreten. Dadurch kommt es zum Zellzyklus-Arrest in der mitotischen G2-Phase, wodurch die Zellteilung der eukaryotischen Zellen gehemmt wird. Da die Zellen weiter wachsen ohne sich zu teilen, kommt es zur sogenannten Megalozytose, welche oft zum Zelltod führt. Im Jahr 2006 wurde erstmals die *pks*-Insel als dafür verantwortliche PAI beschrieben. Diese Insel besitzt eine Sequenzlänge von ca. 54,5 kb und synthetisiert das Protein Colibaktin, welches die beschriebenen Effekte auf eukaryotische Zellen ausübt. Die *pks*-Insel ist in der Art *E. coli* nur

in Stämmen der phylogenetischen Gruppe B2 zu finden, in deren Genom sie immer in der *asnW* tRNA inseriert ist (Nougayrede et al., 2006). Außerdem ist sie nur in Stämmen zu finden, die auch die HPI besitzen. Die schematische Darstellung der PAI ist in Abbildung 7 zu sehen. Die *pks*-Insel ist von zwei DRs flankiert und besitzt ein Integrase-Gen, jedoch keine Gene für die Mobilisierung und den Transfer der Insel. Zwischen den Genen *clbB* und *clbR* ist eine Tandem-Region (VNTR; *variable numbers of tandem repeats*), bestehend aus 8 bp, mit variabler Anzahl (zwischen zwei und 20) an Wiederholungen zu finden. Die Funktion dieser Region konnte noch nicht geklärt werden (Putze et al., 2009). Die *pks*-Insel wurde bereits komplett sequenziert, und die Sequenz ist öffentlich zugänglich (Nougayrede et al., 2006). Die PAI ist auch in anderen Mitgliedern der Familie *Enterobacteriaceae* zu finden, allerdings innerhalb der diversen Region III einer ICE-Typ-HPI. Dies ist für Stämme der Arten *K. pneumoniae*, *C. koseri* und *E. aerogenes* berichtet worden. Dieser ICE-Typ wurde auch in *E. coli*-Stämmen der phylogenetischen Gruppe B1 gefunden, die die herkömmliche Variante der *pks*-Insel nicht besitzen (Putze et al., 2009).

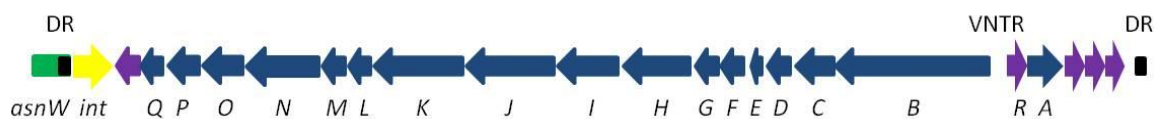


Abbildung 7: Schematische Darstellung der *pks*-Insel. Die PAI wird von zwei DRs (Schwarz) flankiert und ist in die *asnW* tRNA (Grün) inseriert. Der für die Colibaktin-Synthese verantwortliche Gen-Cluster ist in Blau dargestellt. Die Tandem-Region (VNTR) ist zwischen den Genen *clbB* und *clbR* lokalisiert.

4.3. *serU*-Insel

Eine ebenfalls erst kürzlich in *E. coli* entdeckte PAI ist die *serU*-Insel. Diese hat eine Sequenzlänge von ca. 27 kb und synthetisiert das Protein TcpC. Es wurde gezeigt, dass dieser Virulenzfaktor mit der angeborenen Immunantwort des Wirts direkt interagiert und die NF- κ B-Signalkaskade und somit die proinflammatorische Zytokin-Antwort hemmt. TcpC hat eine gewisse Homologie zu den Motiven der TIR-Domäne von Toll-like-Rezeptoren und kann diese somit inhibieren. Tcp steht für *TIR domain containing protein* (TIR-Domänen-enhaltende-Proteine) (Cirl et al., 2008). Die Insel ist innerhalb der Art *E. coli* nur in Stämmen der phylogenetischen Gruppe B2 und in Begleitung der HPI zu finden. Die PAI ist immer in der *serU* tRNA inseriert, wird von zwei DRs flankiert und besitzt keine Gene für die Mobilisierung und den Transfer der Insel. Die schematische Darstellung der PAI ist in

Abbildung 8 zu sehen. Es konnte gezeigt werden, dass die *serU*-Insel zusammen mit der HPI *en bloc* mittels HGT übertragen wurde. Da die *serU*-Insel hoch konserviert ist, wird davon ausgegangen, dass sie vor noch nicht zu langer Zeit in die Art *E. coli* übertragen wurde (Schubert et al., 2010). Es wurden ebenfalls Tcps von anderen bakteriellen Arten, wie z.B. TcpYI von *Yersinia pseudotuberculosis*, beschrieben. TcpYI kann, genau wie TcpC von *E. coli*, mit dem Immunsystem des Wirts interagieren und die Bakterien können dieses somit umgehen (Norenberg et al., 2013).

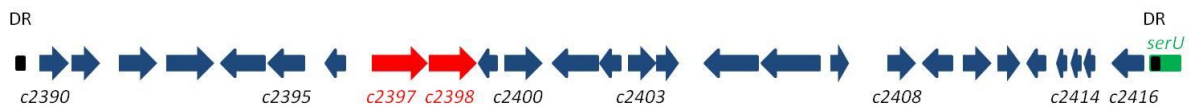


Abbildung 8: Schematische Darstellung der *serU*-Insel. Gezeigt werden die Gene *c2390* bis *c2416*, allerdings sind zur besseren Darstellung nicht alle Gene gekennzeichnet. Die PAI ist in die *serU* tRNA (Grün) inseriert und von zwei DRs flankiert. Die Gene *c2397* und *c2398* (Rot) sind für die Synthese des TcpC-Proteins verantwortlich.

4.4. PAI III

Die PAI III codiert für die Synthese des Siderophors Salmochelin, welches seinen Namen daher erhielt, dass der Rezeptor IroN zuerst in einem *Salmonella enterica*-Stamm identifiziert wurde (Baumler et al., 1998). Später wurde dieser Rezeptor auch in ExPEC-Stämmen gefunden (Russo et al., 1999). IroN ist auf dem *iroBCDEN* Gen-Cluster codiert und an der Aufnahme von Salmochelin, aber auch von anderen Siderophoren, beteiligt. Eine schematische Darstellung der PAI III ist in Abbildung 9 zu sehen. IroN wird häufig in virulenten ExPEC-Stämmen gefunden und besitzt eine erhöhte Expression bei Wachstum im Urin (Bauer et al., 2002). Es wird vermutet, dass IroN bei weiteren virulenten Eigenschaften, wie der Anheftung an Zellen des Harntrakts, beteiligt ist. Da der Harntrakt eine eisenarme Umgebung darstellt, spielt die PAI III besonders bei der Verursachung von Harnwegsinfektionen eine wichtige Rolle (Russo et al., 2002; Russo et al., 2003; Negre et al., 2004; Barasch und Mori, 2004). Wie bereits erwähnt, wurde der *iroBCDN* Gen-Cluster auf dem Plasmid pLVPK von *K. pneumoniae* identifiziert (Chen et al., 2004). Ein *iroBCDEN* Gen-Cluster wurde auch auf dem transferierbaren Plasmid p300 eines ExPEC Stammes gefunden, das Ähnlichkeiten zu dem Plasmid R64 von *S. enterica* besitzt. Dieser Cluster besitzt eine hohe Homologie zu dem chromosomalen *iroBCDEN* Gen-Cluster von *E. coli*

(Sorsa et al., 2003). Es wurden innerhalb der PAI III einige Überreste von IS-Elementen und Transposons gefunden (Dobrindt et al., 2002).

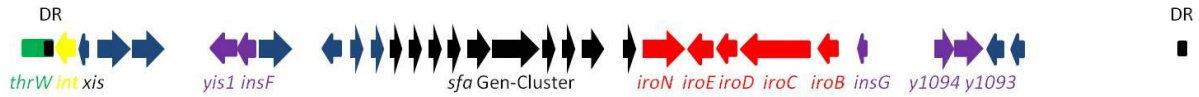


Abbildung 9: Schematische Darstellung der PAI III. Die Insel ist in der *thrW* tRNA (Grün) inseriert und wird von zwei DRs flankiert. Die Integrase ist in Gelb dargestellt. Gezeigt werden der *sfa*- (Schwarz) und der *iro*- (Rot) Gen-Cluster. Überreste von Transposasen und IS-Elemente sind in Lila dargestellt.

5. Next-Generation Sequencing

Der Zusammenbau der ersten menschlichen Genom-Sequenz dauerte 12 Jahre und kostete annähernd drei Milliarden US-Dollar. Bis zur Vollendung des humanen Genom-Projekts im Jahre 2002 waren hunderte Wissenschaftler an dieser Aufgabe beteiligt. Das Projekt wurde mit der sogenannten Sanger-Sequenzierung durchgeführt, die als erste Generation der Sequenzierung bezeichnet wird (Muzzey et al., 2015). Diese DNA-Sequenzierungstechnologie wurde 1977 von Frederick Sanger erstmals vorgestellt, für die er 1980 den Nobelpreis für Chemie erhielt (Liu et al., 2012). Weniger als 15 Jahre nach der Entschlüsselung des ersten menschlichen Genoms kann ein gesamtes humanes Genom in nur einem Tag für ca. 1.000 US-Dollar sequenziert werden (Muzzey et al., 2015). Dies ist möglich durch Sequenzier-Einheiten der nächsten Generation, da diese durch parallele Sequenzierung einen sehr hohen Durchsatz haben (Luthra et al., 2015).

Der Begriff *Next-Generation Sequencing* (NGS) beschreibt nicht eine einzelne Technik. In dem letzten Jahrzehnt wurden Sequenzierer von Firmen entwickelt, die verschiedene Technologien verwenden, wie z.B. Sequenzierung durch Synthese (Illumina), Sequenzierung durch Ligation (Life Technologies) und Ionen-Halbleiter-Sequenzierung (Thermo Fisher Scientific) (Muzzey et al., 2015). Allerdings sind bei NGS immer dieselben Schritte notwendig: Vorbereitung der Proben, Sequenzierung und Darstellung sowie die Auswertung der Rohdaten (Metzker, 2010).

Da für diese Arbeit ein Illumina HiSeq 2000 Sequenzierer verwendet wurde, der die Technologie "Sequenzierung durch Synthese" nutzt, wird diese im Folgenden behandelt und

ist schematisch in Abbildung 10 zu sehen. Als erster Schritt wird genomische DNA (gDNA) von einem Organismus isoliert. Nach einem enzymatischen Verdau oder einer mechanischen Fragmentierung werden anschließend sogenannte Adapter an die Enden der gDNA gebunden. Dadurch können Millionen von verschiedenen einzelsträngigen DNA-Abschnitten an eine Glasoberfläche angebracht und später vervielfältigt werden. Um im nächsten Schritt DNA-Sequenzen sichtbar zu machen, werden Fluoreszenz-markierte Basen verwendet, die nicht durch nachfolgende Basen erweiterbar sind. Es gibt vier verschiedene Farben für die vier Basen Adenin (A), Thymin (T), Guanin (G) und Cytosin (C). Jeder komplementäre Strang eines einst einzelsträngigen DNA-Abschnitts wird nach Zugabe der markierten Basen durch ein Polymerase-Enzym um eine Base verlängert, was zu einem entsprechenden Lichtsignal führt. Ein Mikroskop nimmt die Position des DNA-Abschnitts und seine fluoreszierende Farbe auf. Durch die Auswaschung der übrig gebliebenen markierten Basen verlieren auch die neu eingefügten ihre Fluoreszenz und sind nun erweiterbar. Dieser Vorgang wird solange wiederholt, bis alle komplementären Stränge aufgefüllt und somit die Sequenz der DNA-Abschnitte entschlüsselt wurden (Muzzey et al., 2015).

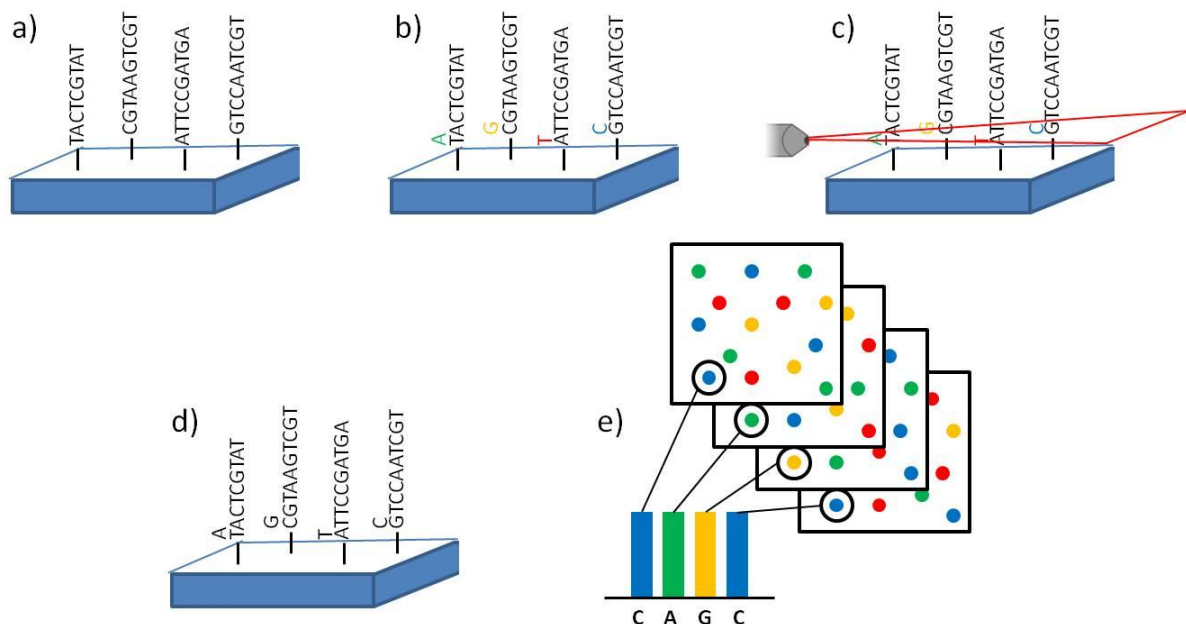


Abbildung 10: Schematische Darstellung des Ablaufs der von Illumina entwickelten Technologie "Sequenzierung durch Synthese". a) Nach der Isolation und dem Verdau der gDNA wird diese mit Hilfe von Adaptern an eine Oberfläche angebracht. Die gDNA ist hier einzelsträngig. b) Die zugegebenen Floreszenz-markierten Basen binden komplementär an den Einzelstrang. c) Mit Hilfe eines Mikroskops wird die Position und die Farbe der fluoreszierenden Base aufgenommen. d) Durch die Auswaschung verlieren die neu eingebauten Basen ihre Fluoreszenz und werden erweiterbar. Die Schritte b) bis d) werden sooft wiederholt, bis alle komplementären Stränge aufgefüllt sind. e) Anhand der aufgenommenen Farben der fluoreszierenden Basen ist die Sequenzabfolge aller DNA-Abschnitte bekannt. Diese werden als Rohdaten bezeichnet.

Im letzten Schritt werden die Ergebnisse *in silico* unter Benutzung einer entsprechenden Software vorbereitet. Zuerst muss die Qualität der Rohdaten überprüft werden. Hier spielt der sogenannte Phred-Qualitätswert eine wichtige Rolle. Dieser beschreibt die Wahrscheinlichkeit einer falsch eingebauten Base in einen komplementären Strang. Je höher der Phred-Wert, desto besser ist die Qualität der erstellten Sequenzen. Dies ist auch in Tabelle 1 zu sehen. Ein weiterer Aspekt ist die Eindeutigkeit der eingebauten Basen. Dieser Wert wird in Prozent angegeben, wobei 100% eine komplett eindeutige Sequenz bedeutet (Ewing et al., 1998; Ewing und Green, 1998). Bevor es letztendlich zur Auswertung der Genome kommen kann, müssen die kurzen DNA-Abschnitte zu längeren Sequenzen assembliert werden. Dafür gibt es verschiedene Möglichkeiten wie die Assemblierung an eine Referenz und die *de novo* Assemblierung (Lee et al., 2015; Arun-Chinnappa und McCurdy, 2015). Mit den fertigen Sequenzen können dann verschiedene Vergleiche angefertigt werden. Eine sehr gängige Methode, um die Verwandtschaft von Bakterien zu untersuchen, ist die Erstellung von phylogenetischen Stammbäumen. Hierfür können unterschiedliche Algorithmen zur Berechnung dieser Bäume verwendet werden (Manzoor et al., 2015). Ein häufig genutzter Algorithmus ist *Neighbor-Joining*, der auch für die hier durchgeführten Untersuchungen verwendet wurde (Lyubetsky et al., 2016).

Tabelle 1: Erläuterung des Phred-Qualitätswerts anhand der Wahrscheinlichkeit einer falsch eingebauten Base und der Genauigkeit der eingebauten Basen.

Phred-Qualitätswert	Wahrscheinlichkeit einer falsch eingebauten Base	Genauigkeit der eingebauten Basen
10	1 von 10	90%
20	1 von 100	99%
30	1 von 1.000	99,9%
40	1 von 10.000	99,99%
50	1 von 100.000	99,999%
60	1 von 1.000.000	99,9999%

6. Ziele dieser Arbeit

In der vorliegenden Arbeit wird der HGT von verschiedenen PAIs innerhalb der Familie *Enterobacteriaceae* untersucht. Insbesondere die Frage, wie sich passiv transferierbare Inseln innerhalb der ExPEC verbreitet haben, ist noch weitgehend ungeklärt. Der Transfer von ICE-Typ-PAIs, die für die Weitergabe entsprechende Gene besitzen, ist bisher recht gut untersucht worden. Da aber innerhalb der ExPEC größtenteils PAIs bekannt sind, die in der Regel nicht selbst übertragbar sind, sollte diese Arbeit vor allem tiefere Einblicke in deren Transfermechanismen liefern.

Dafür wurden zwei Strategien gewählt, um den HGT der PAIs von verschiedenen Seiten zu durchleuchten. Der Großteil dieser Arbeit befasste sich mit der Entschlüsselung ganzer Genome durch NGS und der bioinformatischen Auswertung der erstellten Sequenzen, mit Fokus auf die Inseln. Zudem sollten die PAI-Transfers *in vitro* nachgestellt und mit den *in silico*-Daten verglichen werden.

Es sollte auch gezeigt werden, dass die mittels NGS gewonnenen *in silico*-Daten für unterschiedliche Untersuchungen genutzt werden können. Hierzu zählt neben den phylogenetischen Studien der PAIs und dem angrenzenden *Backbone*-Genom auch die Analyse der für MLST-Schemata verwendeten *Housekeeping*-Gene. Insbesondere wird hier untersucht, ob mehr genetische Informationen eines Stammes zu einer veränderten Darstellung der phylogenetischen Stammbäume führen. Bezüglich der *in vitro*-Transfers mit einem F'-Plasmid sollte durch die Sequenzierung der Transkonjuganten das Ausmaß der übertragenen DNA analysiert werden.

B. Material und Methoden

1. Geräte

Die verwendeten Geräte und deren Hersteller mit Firmensitz sind in Tabelle 2 alphabetisch aufgelistet.

Tabelle 2: Auflistung der verwendeten Geräte.

Gerätetyp bzw. Software	Gerätename und Hersteller
Autoklaven	Laboklav 160-MSLV; SHP Steriltechnik (Detzel, Deutschland) VARIOKLAV Dampfsterilisator Typ 500EC; Thermo Electron LED GmbH (Ismaning, Deutschland)
Bunsenbrenner	Gasprofi 2SCS; WLD Tec (Göttingen, Deutschland)
Datenverarbeitung	MS Office 2010; Microsoft (Redmond, WA, USA)
DNA Analyse Software	CLC Genomics Workbench 6.5, CLC bio (Aarhus, Dänemark)
Elektrophorese-Systeme	PerfectBlue Gelsystem Mini L; Peqlab (Erlangen, Deutschland) Mini-Protean II Electrophoresis System; Bio-Rad (Hercules, CA, USA)
Elektroporator	Gene Pulser II; Bio-Rad (Hercules, CA, USA) Pulse Controller II; Bio-Rad (Hercules, CA, USA)
Gel-Dokumentation	GDS 3.39; Intas (Göttingen, Deutschland)
Inkubatoren	Heraeus T12; Heraeus (Hanau, Deutschland) Cytoperm2; Heraeus (Hanau, Deutschland) Binder BT 115; Binder WTC (Tuttlingen, Deutschland)
Heizblock	Thermomixer comfort; Eppendorf (Hamburg, Deutschland)
Magnetrührer	RCT-basic; IKA-Werke (Staufen im Breisgau, Deutschland)
MALDI-TOF	microflex LT; Bruker (Bremen, Deutschland)
Mikropipetten	Eppendorf (Hamburg, Deutschland)
Mikroplattenleser	Tristar LB 941; Berthold (Bad Wildbach, Deutschland)
Mikrowelle	M1712N; Samsung (Seoul, Südkorea)
NanoDrop	NanoDrop 2000; Thermo Fisher Scientific (Schwerte, Deutschland)
PCR-Thermocycler	GeneAmp PCR System 9700; Applied Biosystems (Thermo Fisher Scientific; Schwerte, Deutschland) Mastercycler gradient; Eppendorf (Hamburg, Deutschland)
Phoenix	Phoenix 100; Becton Dickinson (Heidelberg, Deutschland)
Fotodrucker	P93D; Mitsubishi (Minato, Japan)
Photometer	Ultrospec 3100; Amersham Biosciences (General Electric; Fairfield, CT, USA)
pH-Meter	PB-11; Sartorius (Göttingen, Deutschland)
Schüttler	Thermomixer comfort; Eppendorf (Hamburg, Deutschland)
Schüttelinkubatoren	Certomat R + Certomat H; Sartorius (Göttingen, Deutschland)

	Multitron; INFORS HT (Einsbach, Deutschland)
Sequenzierer	HiSeq 2000; Illumina (San Diego, CA, USA)
Speedvac	SC110; Savant (Holbrook, NY, USA)
Sterilbank	HERA safe KS12; Thermo Electron Corporation (Thermo Fisher Scientific; Schwerte, Deutschland)
Stromversorger	Power Pac 200; Biorad (Hercules, CA, USA) Power Pac HC; Biorad (Hercules, CA, USA)
UV-Transiluminator	MG 21 x 26 UV; Intas (Göttingen, Deutschland)
Vortex-Mischer	MS2 Minishaker; IKA-Werke (Staufen im Breisgau, Deutschland)
Zentrifugen	5810R; Eppendorf (Hamburg, Deutschland) 5417C; Eppendorf (Hamburg, Deutschland) 5417R; Eppendorf (Hamburg, Deutschland) 3K30; Sigma (Osterode am Harz, Deutschland) Sorvall Super T21, Thermo Electron Corporation (Thermo Fisher Scientific; Schwerte, Deutschland)

2. Chemikalien, Enzyme, Puffer und kommerzielle Labor-Kits

Alle verwendeten Chemikalien wurden bei den Firmen Carl-Roth (Karlsruhe, Deutschland), Merck (Darmstadt, Deutschland), Oxoid (Wesel, Deutschland) und Sigma-Aldrich (St. Louis, MO, USA) bezogen.

Die verwendeten Restriktionsenzyme wurden bei den Firmen New England Biolabs (Frankfurt am Main, Deutschland) und Thermo Fisher Scientific (Schwerte, Deutschland) gekauft. Die Ligasen und Polymerasen wurden von den Firmen Thermo Fisher Scientific (Schwerte, Deutschland) und MP Biomedicals (Eschwege, Deutschland) erworben.

Puffer:

a) TAE-Puffer (Tris-Acetat-EDTA, 50-fach konzentriert):

2 M	Tris
6% (v/v)	Essigsäure (99,7%)
50 mM	EDTA (pH 8,0)
ad 1 l H ₂ O _{dest.}	

b) PBS-Puffer (Phosphatgepufferte Salzlösung, 10-fach konzentriert):

80 g NaCl
2 g KCl
17,8 g Na₂HPO₄ – 2 H₂O
2,4 g KH₂PO₄
ad 1 l H₂O_{dest.}, pH-Wert 7,4

Die in Tabelle 3 aufgeführten Labor-Kits wurden für diese Arbeit verwendet.

Tabelle 3: Auflistung der kommerziell zu erwerbenden Labor-Kits für diese Arbeit.

Name	Kat. #	Hersteller
<i>High Pure PCR Template Preparation Kit</i>	1796828	Roche Diagnostics, Unterhaching, Deutschland
<i>QIAprep Spin Miniprep Kit</i>	27106	Qiagen, Hilden, Deutschland
<i>QIAquick PCR Purification Kit</i>	28106	Qiagen, Hilden, Deutschland
<i>Nucleo Bond[®] PC 20 Kit</i>	740571.100	Macherey Nagel, Düren, Deutschland
<i>Taq DNA Core Kit</i>	EPTQK109	MP Biomedicals, Eschwege, Deutschland
<i>Ultra Clean Gel Spin – DNA Extraction Kit</i>	12400-250	MO Bio, Carlsbad, CA, USA

3. Nährmedien

Zur Kultivierung der Bakterienstämme wurden verschiedene Flüssig- und Festmedien verwendet. Vor der Verwendung wurden diese bei 121°C für 20 min autoklaviert. Zur Selektion wurden je nach Bakterienstamm verschiedene Antibiotika hinzugegeben.

3.1. Flüssigmedien

Die Flüssigmedien wurden vor allem zur Kultivierung von üN-Kulturen verwendet, die für verschiedene Experimente erforderlich waren. Die Bakterien befinden sich in diesem Fall in der stationären Wachstumsphase.

a) LB (*Lysogeny Broth*) -Medium:

10 g Trypton
5 g Hefeextrakt
5 g NaCl
ad 1 l H₂O_{dest}, pH-Wert 7,5

b) SOC (*Super Optimal broth with Catabolite repression*) -Medium:

20 g Trypton
5 g Hefeextrakt
0,5 g NaCl
10 ml KCl-Lösung (250 mM)
ad 1 l H₂O_{dest}, pH-Wert 7,0

c) MH (Müller-Hinton) -Medium:

21 g Müller-Hinton-Bouillon (Oxoid; Wesel, Deutschland; Kat. # CM0337)
ad 1 l H₂O_{dest}, pH-Wert 7,3

d) MM9 Minimal-Medium (10-fach konzentriert):

0,3 g KH₂PO₄
0,5 g NaCl
1,0 g NH₄Cl
ad 100 ml H₂O_{dest}

3.2. Festmedien

Zur Erstellung von Festmedien, auch als Agarplatten bekannt, wurden den Flüssigmedien Agar (Becton Dickinson, Heidelberg, Deutschland; Kat. # 214030) hinzugegeben. Durch Erhitzen auf 121°C wurde dies sterilisiert, dann langsam abgekühlt, um bei ca. 50°C die benötigten Antibiotika hinzugeben. Anschließend wurde der Agar in Petrischalen gegossen.

Für Petrischalen der Größe 94 mm x 16 mm wurden jeweils ca. 20 ml benötigt, für die Größe 145 mm x 20 mm ca. 50 ml. Nach Erkalten der Platten wurden diese bei 4°C aufbewahrt.

a) LB-Agarplatten:

LB-Medium mit 1,5% (w/v) Agar

b) Columbia Blut-Agarplatten (Becton Dickinson, Heidelberg, Deutschland; Kat. # 254071):

LB-Agar mit 5% (v/v) Schaf-Erythrozyten

c) *Yersinia*-Selektivagar

58 g *Yersinia*-Selektivagar, Kat. # 95760-500G (Fluka, Sigma-Aldrich Chemie; Buchs, Schweiz)

ad 1 l H₂O_{dest}, pH-Wert 7,4

d) CAS (Chrom Azurol S) -Agar

PIPES-Agar (50°C)	170 ml
10x LB	6 ml
1 M MgSO ₄	0,4 ml
1 M Na ₂ SO ₄	0,4 ml
0,1 M CaCl ₂	0,2 ml
20% Glucose	2 ml
CAS-Lösung	20 ml

CAS-Lösung:

Lösung 1) 121 mg Chrom Azurol S in 100 ml H₂O_{dest}
 + 20 ml von 1 mM FeCl₃ (gelöst in 10 mM HCl)

Lösung 2) 145,8 mg HDTMA in 80 ml H₂O_{dest}

Lösung 1) langsam in Lösung 2) rühren und autoklavieren

PIPES -Agar:

100 ml MM9 Medium (10-fach konzentriert)

+ 15 g Agar

ad 500 ml H₂O_{dest}

unter Rühren 31,1 g PIPES (Piperazindiethansulfonsäure) hinzugeben

pH mit NaOH auf 6,8 einstellen

ad 850 ml H₂O_{dest}

4. Antibiotika

Die verwendeten Antibiotika wurden dafür genutzt, um Bakterienstämme oder Transkonjuganten zu selektieren und um sicherzustellen, dass manche Stämme ihr tragendes Plasmid nicht verlieren. Die Antibiotika wurden den Nährmedien in den in Tabelle 4 aufgeführten Konzentrationen hinzugefügt. Die erstellten Stammlösungen wurden bei -20°C gelagert, angebrochene Lösungen wurden für eine befristete Zeit bei 4°C aufbewahrt.

Tabelle 4: Verwendete Antibiotika, deren Lösungsmittel und Konzentrationen.

Antibiotika	Stammlösung	Lösungsmittel	Endkonzentration
Ampicillin	100 mg/ml	H ₂ O _{dest}	100 µg/ml
Chloramphenicol	50 mg/ml	MeOH	30 µg/ml
Kanamycin	50 mg/ml	H ₂ O _{dest}	50 µg/ml
Nalidixinsäure	6 mg/ml	NaOH (0,5 N)	60 µg/ml
Streptomycin	30 mg/ml	H ₂ O _{dest}	30 µg/ml
Tetracyclin	5 mg/ml	70% EtOH	12 µg/ml

Zur Selektion von *Yersinia enterocolitica* wurde das *Yersinia Selective Supplement* (Fluka, Sigma-Aldrich Chemie; Buchs, Schweiz, Kat. # 75258), dem Nährmedium hinzugefügt. In

einem Fläschchen, das für 500 ml Medium oder Agar verwendet wurde, waren 7,5 mg Cefsulodin, 2 mg Triclosan (Irgasan) und 1,25 mg Novobiocin enthalten.

Um zu überprüfen, welche Antibiotika-Resistenzen bei gewissen Stämmen auftreten, wurde der sogenannte Agardiffusionstest angewendet. Dafür wurden die Bakterien auf einer Agarplatte ausgestrichen und mit kleinen Plättchen, die Antibiotika enthalten, versehen. Nach Kultivierung üN war es möglich zu sehen, ob Antibiotika-bedingte Wachstumsdefizite, so genannte Hemmhöfe, auftreten. Die verwendeten Antibiotika-Plättchen waren von der Firma Oxoid (Wesel, Deutschland). Die Antibiotika und deren Konzentrationen pro Plättchen (Beschickung) sind in Tabelle 5 zu finden.

Tabelle 5: Verwendete Antibiotika-Plättchen.

Antibiotikum	Beschickung
Ampicillin	10 µg
Chloramphenicol	30 µg
Kanamycin	30 µg
Nalidixinsäure	30 µg
Rifampicin	2 µg
Streptomycin	25 µg
Tetracyclin	30 µg
Trimethoprim	5 µg

5. Bakterienstämme

Für die unterschiedlichen Projekte wurden verschiedene Bakterienstämme verwendet, die in den nachfolgenden Tabellen aufgelistet sind. In Tabelle 6 sind die Stämme für die phylogenetische Auswertung der Pathogenitätsinseln zu finden, in Tabelle 7 die für den Transfer der immobilisierenden Inseln mit einem F'-Plasmid. Alle Stämme, die die Konjugations-Versuche mit der mobilen PAI betreffen, sind in Tabelle 8 aufgeführt.

Die *Escherichia coli* Referenz (ECOR)-Sammlung besteht aus 72 Stämmen und wurde für die Erstellung der phylogenetischen Stammbäume verwendet. In Tabelle 6 ist die Herkunft und der Wirt, aus dem das Bakterium isoliert wurde, aufgelistet (Ochman und Selander, 1984). Zusätzlich wurden die *E. coli*-Stämme 536, S107 und S108 verwendet.

Tabelle 6: Verwendete Bakterienstämme der Familie *Enterobacteriaceae* für die Erstellung der phylogenetischen Bäume.

Bakterienstämme für die Phylogenie der immobilisierten PAIS		
Stamm	Wirt	Herkunft
<i>E. coli</i> ECOR01	Mensch (♀)	IA, USA
<i>E. coli</i> ECOR02	Mensch (♂)	NY, USA
<i>E. coli</i> ECOR03	Hund	MA, USA
<i>E. coli</i> ECOR04	Mensch (♀)	IA, USA
<i>E. coli</i> ECOR05	Mensch (♀)	IA, USA
<i>E. coli</i> ECOR06	Mensch (♂)	IA, USA
<i>E. coli</i> ECOR07	Orang-Utan	WA, USA (Zoo)
<i>E. coli</i> ECOR08	Mensch (♀)	IA, USA
<i>E. coli</i> ECOR09	Mensch (♀)	Schweden
<i>E. coli</i> ECOR10	Mensch (♀)	NY, USA
<i>E. coli</i> ECOR11	Mensch (♀)	Schweden
<i>E. coli</i> ECOR12	Mensch (♀)	Schweden
<i>E. coli</i> ECOR13	Mensch (♀)	Schweden
<i>E. coli</i> ECOR14	Mensch (♀)	Schweden
<i>E. coli</i> ECOR15	Mensch (♀)	Schweden
<i>E. coli</i> ECOR16	Leopard	WA, USA (Zoo)
<i>E. coli</i> ECOR17	Schwein	Indonesien
<i>E. coli</i> ECOR18	Celebes-Affe	WA, USA (Zoo)
<i>E. coli</i> ECOR19	Celebes-Affe	WA, USA (Zoo)
<i>E. coli</i> ECOR20	Ochse	Bali
<i>E. coli</i> ECOR21	Ochse	Bali
<i>E. coli</i> ECOR22	Ochse	Bali
<i>E. coli</i> ECOR23	Elefant	WA, USA (Zoo)
<i>E. coli</i> ECOR24	Mensch (♀)	Schweden
<i>E. coli</i> ECOR25	Hund	NY, USA
<i>E. coli</i> ECOR26	Mensch (Kleinkind)	MA, USA
<i>E. coli</i> ECOR27	Giraffe	WA, USA (Zoo)
<i>E. coli</i> ECOR28	Mensch (♀)	IA, USA
<i>E. coli</i> ECOR29	Känguru-Ratte	NV, USA
<i>E. coli</i> ECOR30	Bison	AB, Kanada
<i>E. coli</i> ECOR31	Leopard	WA, USA (Zoo)
<i>E. coli</i> ECOR32	Giraffe	WA, USA (Zoo)
<i>E. coli</i> ECOR33	Schaf	CA, USA
<i>E. coli</i> ECOR34	Hund	MA, USA
<i>E. coli</i> ECOR35	Mensch (♂)	IA, USA
<i>E. coli</i> ECOR36	Mensch (♀)	IA, USA
<i>E. coli</i> ECOR37	Marmosette	WA, USA (Zoo)
<i>E. coli</i> ECOR38	Mensch (♀)	IA, USA
<i>E. coli</i> ECOR39	Mensch (♀)	Schweden
<i>E. coli</i> ECOR40	Mensch (♀)	Schweden
<i>E. coli</i> ECOR41	Mensch (♂)	Tonga
<i>E. coli</i> ECOR42	Mensch (♂)	MA, USA

<i>E. coli</i> ECOR43	Mensch (♀)	Schweden
<i>E. coli</i> ECOR44	Puma	WA, USA (Zoo)
<i>E. coli</i> ECOR45	Schwein	Indonesien
<i>E. coli</i> ECOR46	Celebes-Affe	WA, USA (Zoo)
<i>E. coli</i> ECOR47	Schaf	Neuguinea
<i>E. coli</i> ECOR48	Mensch (♀)	Schweden
<i>E. coli</i> ECOR49	Mensch (♀)	Schweden
<i>E. coli</i> ECOR50	Mensch (♀)	Schweden
<i>E. coli</i> ECOR51	Mensch (Kleinkind)	MA, USA
<i>E. coli</i> ECOR52	Orang-Utan	WA, USA (Zoo)
<i>E. coli</i> ECOR53	Mensch (♀)	IA, USA
<i>E. coli</i> ECOR54	Mensch	IA, USA
<i>E. coli</i> ECOR55	Mensch (♀)	Schweden
<i>E. coli</i> ECOR56	Mensch (♀)	Schweden
<i>E. coli</i> ECOR57	Gorilla	WA, USA (Zoo)
<i>E. coli</i> ECOR58	Löwe	WA, USA (Zoo)
<i>E. coli</i> ECOR59	Mensch (♂)	MA, USA
<i>E. coli</i> ECOR60	Mensch (♀)	Schweden
<i>E. coli</i> ECOR61	Mensch (♀)	Schweden
<i>E. coli</i> ECOR62	Mensch (♀)	Schweden
<i>E. coli</i> ECOR63	Mensch (♀)	Schweden
<i>E. coli</i> ECOR64	Mensch (♀)	Schweden
<i>E. coli</i> ECOR65	Celebes-Affe	WA, USA (Zoo)
<i>E. coli</i> ECOR66	Celebes-Affe	WA, USA (Zoo)
<i>E. coli</i> ECOR67	Ziege	Indonesien
<i>E. coli</i> ECOR68	Giraffe	WA, USA (Zoo)
<i>E. coli</i> ECOR69	Celebes Affe	WA, USA (Zoo)
<i>E. coli</i> ECOR70	Gorilla	WA, USA (Zoo)
<i>E. coli</i> ECOR71	Mensch (♀)	Schweden
<i>E. coli</i> ECOR72	Mensch (♀)	Schweden
Stamm	relevante Eigenschaften	Referenz
<i>E. coli</i> 536	phyl. Gr. B2, besitzt HPI, <i>pks</i> -Insel und <i>serU</i> -Insel	(Hacker et al., 1992)
<i>E. coli</i> S107	phyl. Gr. B2, besitzt HPI und <i>serU</i> -Insel	(Le et al., 2007)
<i>E. coli</i> S108	phyl. Gr. B2, besitzt HPI und <i>serU</i> -Insel	(Le et al., 2007)
<i>E. coli</i> ED1a	phyl. Gr. B2, besitzt ICE-Typ-HPI, ICE enthält RegX nicht	(Vimont et al., 2012a)
<i>K. pneumoniae</i> HU86	besitzt ICE-Typ-HPI, ICE enthält RegX	Universität Pécs, Ungarn
<i>K. pneumoniae</i> HU145	besitzt ICE-Typ-HPI, ICE enthält RegX	Universität Pécs, Ungarn
<i>K. pneumoniae</i> HU245	besitzt ICE-Typ-HPI, ICE enthält RegX nicht	Universität Pécs, Ungarn
<i>K. pneumoniae</i> B-2495	besitzt ICE-Typ-HPI, ICE enthält RegX	Universität Pécs, Ungarn

<i>K. pneumoniae</i> 1084	besitzt ICE-Typ-HPI, ICE enthält RegX	(Lin et al., 2012)
<i>C. koseri</i> ATCC BAA-895	besitzt ICE-Typ-HPI, ICE enthält RegX	<i>American Type Culture Collection</i>

Es wurden unterschiedliche *E. coli*-Stämme für den Transfer der immobilisierten Inseln mit einem F'-Plasmid verwendet. Hierfür wurden zwei Donoren (NU14 Cm F' und 536 Cm F') und zwei Rezipienten (MG1655 und MG1655 Nal) benutzt. Der Laborstamm XL1-Blue MRF' besaß das benötigte F'-Plasmid. Die Bakterien sind in Tabelle 7 aufgeführt.

Tabelle 7: Verwendete und erstellte *E. coli* Bakterienstämme für die Konjugationen mit dem F'-Plasmid.

Bakterienstämme für den Transfer der immobilisierten PAIs mittels F'-Plasmid		
Stamm	relevante Eigenschaften	Referenz
XL1-Blue MRF'	$\Delta(mcrA)183 \Delta(mcrCB-hsdSMR-mrr)173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac$ [F' <i>proAB lacIqZAM15 Tn10</i> (Tet ^R)]	Stratagene (Agilent; Santa Clara, CA, USA)
NU14	Zystitis-Isolat, phyl. Gr. B2, besitzt HPI und <i>pks</i> -Insel, produziert β -Hämolysin	(Johnson et al., 2001b)
NU14 Cm	chromosomale Insertion einer Chloramphenicol-Resistenzkassette innerhalb der HPI	diese Arbeit
NU14 Cm F'	trägt das F'-Plasmid von XL1-Blue MRF'	diese Arbeit
536	Pyelonephritis-Isolat, phyl. Gr. B2, besitzt HPI, <i>pks</i> -Insel und <i>serU</i> -Insel, produziert β -Hämolysin, Streptomycin-Resistenz	(Hacker et al., 1992)
536 Cm	chromosomale Insertion einer Chloramphenicol-Resistenzkassette innerhalb der HPI	diese Arbeit
536 Cm F'	trägt das F'-Plasmid von XL1-Blue MRF'	diese Arbeit
MG1655	<i>E. coli</i> K-12, -, λ -, <i>ilvG</i> , <i>rfb-50 rph-1</i> , phyl. Gr. A, Streptomycin-Resistenz, produziert kein β -Hämolysin	(Blattner et al., 1997)
MG1655 Nal	spontane Nalidixinsäure-Resistenz	diese Arbeit
MG1655 HPI-Cm	Transkonjugant von NU14 Cm F' und MG1655	diese Arbeit
MG1655 HPI-Cm F'	trägt das F'-Plasmid von NU14 Cm F'	diese Arbeit
MG1655 Nal HPI-Cm	Transkonjugant von 536 Cm F' und	diese Arbeit

	MG1655 Nal	
--	------------	--

Der Transfer der mobilen HPI fand innerhalb der Familie *Enterobacteriaceae* statt. Die verwendeten und erstellten Bakterienstämme sind in Tabelle 8 aufgelistet.

Tabelle 8: Verwendete und erstellte *Enterobacteriaceae* Stämme für den Transfer der mobilen HPI.

Bakterienstämme für den Transfer der mobilen PAI		
Stamm	relevante Eigenschaften	Referenz
HU86	<i>K. pneumoniae</i> , Blutkultur-Isolat, besitzt ICE-Typ-HPI, ICE enthält RegX	Universität Pécs, Ungarn
HU86 Cm	chromosomale Insertion einer Chloramphenicol-Resistenzkassette innerhalb der HPI	diese Arbeit
ED1a	<i>E. coli</i> , Stuhl-Isolat (Kommensale), phyl. Gr. B2, besitzt ICE-Typ-HPI, ICE enthält RegX nicht	(Vimont et al., 2012b)
ED1a Cm	chromosomale Insertion einer Chloramphenicol-Resistenzkassette innerhalb der HPI	diese Arbeit
EDF72	<i>E. coli</i> , Harnwegsinfekt-Isolat, phyl. Gr. B2, besitzt ICE-Typ-HPI, ICE enthält RegX nicht	Straubing
EDF72 Kn	chromosomale Insertion einer Kanamycin-Resistenzkassette innerhalb der HPI	diese Arbeit
MG1655	<i>E. coli</i> K-12, phyl. Gr. A, HPI negativ, Streptomycin-Resistenz	(Blattner et al., 1997)
MRS40	<i>Y. enterocolitica</i> , O9, HPI und CAS negativ, Nalidixinsäure-Resistenz	(Pelludat et al., 2002)
MG1655 ICE-Cm	Transkonjugant von HU86 Cm und MG1655	diese Arbeit
MRS40 ICE-Cm	Transkonjugant von HU86 Cm und MRS40	diese Arbeit

6. Plasmide

Die Plasmide, die für diese Arbeit verwendet wurden, sind in Tabelle 9 aufgelistet. Die Plasmide pKD46, pKD3 und pKD4 wurden für die Insertion einer Antibiotika-Resistenzkassette verwendet, das F'-Plasmid für den Transfer der immobilisierten PAIs.

Tabelle 9: Verwendete Plasmide und deren genetischen Eigenschaften.

Plasmid	Eigenschaften	Referenz
pKD46	<i>repA101</i> (ts), <i>araBp-gam-bet-exo</i> (λ Red Rekombinase unter Kontrolle eines <i>araB</i> Promotors), <i>Ap^R</i> (<i>bla</i>)	(Datsenko und Wanner, 2000)
pKD3	<i>oriRg</i> , <i>Ap^R</i> , <i>cat</i> -Gen (<i>Cm^R</i>) flankiert von FRT Sites	(Datsenko und Wanner, 2000)
pKD4	<i>oriRg</i> , <i>Ap^R</i> , <i>npt</i> -Gen (<i>Kn^R</i>) flankiert von FRT Sites	(Datsenko und Wanner, 2000)
F' Episom	<i>proAB lacI^fZAM15 Tn10</i> (<i>Tet^R</i>)	Stratagene (Agilent; Santa Clara, CA, USA)

7. Oligonukleotide

Zur Durchführung von Polymerase-Kettenreaktionen (PCRs) und Sequenzierungen wurden Oligonukleotide, sogenannte Primer, benötigt. Diese wurden bei der Firma Metabion (Planegg, Deutschland) erworben. Alle in dieser Arbeit verwendeten Oligonukleotide sind in Tabelle 10 aufgelistet. Die Sequenz ist in der Leserichtung der Polymerase, vom 5' Anfang zum 3' Ende, angegeben.

Tabelle 10: Verwendete Oligonukleotide, deren Sequenz und Verwendungszweck.

Primer	Sequenz (5' → 3')	Verwendung
ChuA.1	GACGAACCAACGGTCAGGAT	phylogenetische Gruppe
ChuA.2	TGCCGCCAGTACCAAAGACA	phylogenetische Gruppe
YjaA.1	TGAAGTGTCAGGAGACGCTG	phylogenetische Gruppe
YjaA.2	ATGGAGAATGCGTTCCTCAAC	phylogenetische Gruppe
TspE4C2.1	GAGTAATGTCGGGGCATTCA	phylogenetische Gruppe
TspE4C2.2	CGCGCCAACAAAGTATTACG	phylogenetische Gruppe
fyuA.1080.for	CTACGACATGCCGACAATGCC	HPI
fyuA.1709.rev	TGCTTCCC CGCCATAACGTG	HPI
clbA.IHE.for	TAACTTCCTTCACTATCTCA	<i>pks</i> -Insel
clbA.IHE.rev	GAGAGGCTAATGCGAGAAAT	<i>pks</i> -Insel
clbN.IHE.for	CATCAGCTGCCGAGCGCGCA	<i>pks</i> -Insel
clbN.IHE.rev	ACCTCGTACGCGCGCTTGGC	<i>pks</i> -Insel
tcpC.for	GGCAACAATATGTATAATATCCT	<i>serU</i> -Insel
tcpC.rev	GCCAGTCTATTTCTGCTAAAGA	<i>serU</i> -Insel
iroN.28.for	CTAACTGTGCTCCTGGTTGGGTGA	PAI III
iroN.2053.rev	TGACGCCGACATTAAGACGCAGATT	PAI III
RegX-521.for	TTGGGCTGACTGGTATTAGTGGC	Region X
RegX-358.rev	CGAGTTTTTCCACCGCATCAACG	Region X

mobB.HU86.for	TGAACGGGAATTGCGCACCG	<i>mobB</i>
mobB.HU86.rev	TCCTGTCGGCGCATTGATGG	<i>mobB</i>
INV 862 for	AACTTACAAATGGATTATCGCC	<i>Y. enterocolitica</i>
INV 1535 rev	CTTATCACACCCTGTGCATC	<i>Y. enterocolitica</i>
P1-fyuA-2104.for	TTAATCCATAGTGATTATTTAAACAATGAA ATTGCGATTAGGACAAATAGATTTGTGTA GGCTGGAGCTGCTTC	Insertion einer Antibiotika-Resistenzkassette
P2-fyuA-2691.rev	TCCCAGCGTGGTTTGAGTAGTGTACCTGA GTGATATTTGTGTTATGTATGCATTCATAT GAATATCCTCCTTA	Insertion einer Antibiotika-Resistenzkassette
p34S-Cm2.484.rev	TCACCGTAACACGCCACATCTT	Kontrolle Cm-Kassette
HPI-fyuA-2132.for	AAATTGCGATTAGGACAAATAG	Kontrolle Cm-Kassette
HPI-fyuA-2947.rev	CAACTGCTTCCGTTATAGTGAC	Kontrolle Cm/Kn-Kassette
HPI 30784.for	CGGGAAGCAGCGTGAACGGTGTG	Kontrolle Kn-Kassette
pKD-kt	CGGCCACAGTCGATGAATCC	Kontrolle Kn-Kassette
12kb.pks.for	GCTGTTGGTTGTGCTTTATG	Sequenzierung 12 kb <i>downstream pks</i> -Insel
12kb.pks.rev	CTGATTATTGTGCCAGGACG	Sequenzierung 12 kb <i>downstream pks</i> -Insel
insert.ICE.rev	TAAAGGGTTTATCAGAAGGC	<i>asn</i> tRNA Insertion des ICE (<i>intB1</i> und <i>intB2</i>)
Ec.HPI- border.for.1998	TTTATTCCGTTGCGTGAGGTT	<i>asnT</i> tRNA Insertion des ICE in <i>E. coli</i>
attL-EC31-2633- for	CCGCCATTACTTACAACCAGATT	<i>asnV</i> tRNA Insertion des ICE in <i>E. coli</i>
<i>asnW</i> -U.25544.for	TCTCGCGCCAGAACAACACTTTG	<i>asnU</i> tRNA Insertion des ICE in <i>E. coli</i>
Ec.rev.16500	CAGATAGCGCAAGCAGAGAT	<i>asnW</i> tRNA Insertion des ICE in <i>E. coli</i>
insert.Yent.t054.for	CGTTGGGAAAACGAGAGTCA	<i>asn</i> (t054) tRNA Insertion des ICE in <i>Y. enterocolitica</i>
insert.HU86.for	ATATGCAAGACTTCAGCGGC	<i>asn</i> tRNA Insertion des ICE in <i>K. pneumoniae</i>

8. Software

Die Auswertungen der Versuche und deren graphischen Darstellungen wurden mit dem Software-Paket MS Office (Microsoft; Redmont, WA, USA) durchgeführt. Die Daten der Sequenzierungen wurden mit der CLC Genomics Workbench 6.5 (CLC bio; Aarhus, Dänemark) bearbeitet und graphisch dargestellt. Auch wurden die Versuche zur Insertion einer Antibiotika-Resistenzkassette mit diesem Programm *in silico* geplant. Bereits veröffentlichte Genome und Gene wurden der Internetseite <http://www.ncbi.nlm.nih.gov>

(Stand 06/2015) entnommen. Um *Contigs* aus den Sequenzierungen zu identifizieren, wurde ein sogenanntes BLAST (*Basic Local Alignment Search Tool*) der Internetseite <http://blast.ncbi.nlm.nih.gov> (Stand 06/2015) verwendet. Die verwendeten Parameter waren: “blastn”, “*nucleotide collection (nr/nt)*” und “*highly similar sequences (megablast)*”.

9. Mikrobiologische Methoden

Die hier aufgeführten Methoden befassen sich mit der Handhabung von Bakterien und der Aufnahme von fremder DNA, entweder passiv aus dem Umfeld oder aktiv durch Übertragung eines anderen Bakteriums.

9.1. Allgemeine Kultivierung von Bakterien

Je nach dem, für welche Zwecke die Bakterien gebraucht wurden, wurden diese entweder in Flüssigmedien oder auf Festmedien μ N kultiviert. Dabei handelte es sich um LB-Medium, wenn nicht anders angegeben. Flüssigkulturen wurden in einem Schüttelinkubator bei 200 rpm gezüchtet, Bakterien auf Agarplatten in einem Inkubator. *E. coli*-Stämme und *K. pneumoniae*-Stämme wurden grundsätzlich bei 37°C bebrütet, *Y. enterocolitica*-Stämme bei 30°C. Zur Aufbewahrung der Bakterienstämme wurden diese in 25% Glycerol in H₂O_{dest} (v/v) bei -80°C eingefroren.

9.2. Transformation und Selektion

Transformation beschreibt die Fähigkeit eines Bakteriums, ein Plasmid oder lineare DNA aus der Umgebung aufzunehmen. Diese Fähigkeit kann man sich zu Nutze machen, indem man die Permeabilität der bakteriellen Zellwand durch Hitze oder Strom erhöht und somit die Aufnahme begünstigt. Für diese Arbeit wurde die Methode der Elektroporation angewendet. Damit die Bakterien die DNA aufnehmen können, müssen sie erst elektrokompent werden. Hierfür wurden 40 ml einer Flüssigkultur bei einer OD₆₀₀ von 0,5 bis 0,6 durch Zentrifugation (Z! 4300 rpm, 15 min, 4°C) pelletiert. Die Zellen wurden fünfmal jeweils erst in 1 ml eiskaltem 10% Glycerol in H₂O_{dest} (v/v) resuspendiert, 10 s mit einem Vortex-Mischer geschüttelt und anschließend zentrifugiert (Z! 14000 rpm, 1 min, 4°C). Das Zellpellet wurde final in 400 μ l eiskaltem 10% Glycerol in H₂O_{dest} (v/v) gelöst. Die nun elektrokompenten

Bakterien wurden zu je 50 µl aliquotiert und auf Eis gestellt. Die DNA wurde vor der Transformation mit dem *QIAquick PCR Purification Kit* aufgereinigt und einer 20 minütigen Dialyse (Nitrocellulose Filter auf H₂O_{dest}; MF-Membrane Filters, Filter Typ 0,025 µm VSWP, Millipore (Merck; Darmstadt, Deutschland), Kat. # VSWP04700) zur Entsalzung unterzogen. Einem Aliquot elektrokompenter Bakterien wurden entweder 1 - 5 µl PCR-Produkt oder 1 - 2 µl Plasmid hinzugefügt und für 10 min auf Eis inkubiert, um eine Anlagerung der DNA an die bakterielle Zellwand zu begünstigen. Das Gemisch wurde in eine bei -20°C gelagerte Elektroporationsküvette (Biozym Scientific GmbH; Oldendorf, Deutschland; Kat. # 748010) gegeben. Anschließend wurde die Zellwand durch einen elektrischen Impuls (Widerstand 200 Ω, Ladung 25 µF, Spannung 1,8 kV, Puls 3,7 – 5,1 ms) permeabel für die DNA gemacht. Die Bakterien-Suspension wurde sofort in 1 ml auf 37°C vorgewärmtes SOC-Medium aufgenommen und zur Regeneration der Zellwand in einem Thermomixer bei 37°C und 300 rpm für 1 h inkubiert. Die Bakterien wurden anschließend auf LB-Platten ausplattiert und durch die im Agar enthaltenen Antibiotika bezüglich der Aufnahme der DNA selektiert.

9.3. Konjugation und Selektion

Durch Konjugation kann ein Bakterium über einen Pilus DNA auf ein anderes Bakterium übertragen. In manchen Fällen werden zusätzliche genetische Informationen weiter gegeben (z.B. ICE), in anderen ganze DNA-Abschnitte des Rezipienten durch Teile des Donors ersetzt (z.B. F'-Plasmid). In dieser Arbeit wurde diese Fähigkeit verwendet, um PAIs auf einen Rezipienten zu übertragen. Durch die Markierung einer PAI mit einer Antibiotika-Resistenzkassette konnte der Transfer nachvollzogen und die Transkonjuganten selektiert werden.

9.3.1. Konjugation von Pathogenitätsinseln mit einem F'-Plasmid

Für die Übertragung der immobilisierten PAIs HPI, *pks*-Insel und *serU*-Insel wurde ein F'-Plasmid verwendet, welches die DNA eines Donors auf einen Rezipienten überträgt und ersetzt. Die Herstellung der *E. coli*-Donoren NU14 Cm F' (HPI und *pks*-Insel) und 536 Cm F' (HPI, *pks*-Insel und *serU*-Insel) wurde auf dieselbe Weise durchgeführt. Auch die Konjugationsbedingungen eines zwei-Insel-Transfers und eines drei-Insel-Transfers waren identisch. Die Experimente wurden immer mindestens dreifach durchgeführt. Für die Verdünnungsreihen wurden immer 450 µl PBS in einem 1,5 ml Reaktionsgefäß vorgelegt und

50 µl der vorherigen Verdünnungsstufe hinzugefügt und vermischt. Somit war das Volumen einer Verdünnungsstufe - ausgenommen der letzten - 450 µl.

a) Zuerst wurde die HPI der zukünftigen Donoren mit einer Chloramphenicol-Resistenzkassette nach der Methode von Datsenko und Wanner (Datsenko und Wanner, 2000) markiert (siehe "B.10.6. Insertion von Antibiotika-Resistenzkassetten"). Zum Erhalt dieser Resistenzkassette wurde mit dem Plasmid pKD3 eine PCR mit den Primern P1-fyuA-2104.for und P2-fyuA-2691.rev durchgeführt. Das PCR-Produkt wurde aufgereinigt und dialysiert. Je eine üN-Kultur der WT-Stämme NU14 und 536 wurde elektrokompotent gemacht. Anschließend wurden je 2 µl des Plasmids pKD46 in die kompetenten Zellen transformiert. Die Plasmid-tragenden Stämme wurden üN bei 30°C und 200 rpm inkubiert. Am nächsten Tag wurden je 1 ml der üN-Kultur in 40 ml frisches LB-Medium mit Arabinose (10 mM) und Ampicillin gegeben und bis zu einer OD₆₀₀ von ca. 0,5 bei 30°C und 200 rpm weiter inkubiert. Die Zellen wurden anschließend wieder elektrokompotent gemacht, mit je 5 µl des PCR-Produkts transformiert und in SOC-Medium regeneriert. Die Selektion der Chloramphenicol-resistenten Bakterien erfolgte durch das Ausplattieren auf LB-Agarplatten mit dem genannten Antibiotikum.

b) Zur Übertragung des F'-Plasmids von dem kommerziell erhältlichen Stamm *E. coli* XL1-Blue MRF' auf einen zukünftigen Donor wurde von beiden Stämmen eine üN-Kultur von 5 ml verwendet. Diese wurden vereint und bei 4300 rpm bei RT für 15 min zentrifugiert. Das Zellpellet wurde in 0,5 ml PBS gelöst und auf eine Blut-Agarplatte gegeben. Diese wurde üN bei 37°C kultiviert um das F'-Plasmid zu konjugieren. Am nächsten Tag wurde der angewachsene Bakterienrasen von der Agarplatte abgekratzt und in 1 ml PBS gelöst. Es wurden Verdünnungsreihen von 10⁻¹ bis 10⁻⁵ in PBS erstellt und jeweils 100 µl auf LB-Platten mit Chloramphenicol und Tetracyclin ausplattiert. Nach Inkubation üN bei 37°C wurden die gewachsenen Klone auf das Vorhandensein der HPI, auf deren phylogenetischen Gruppe und auf die Produktion von β-Hämolysin überprüft.

c) Der nun fertige F'-Plasmid-positive Donor und der Rezipient wurden üN in jeweils 10 ml LB-Medium angezüchtet, am folgenden Tag vermischt und pelletiert (Z! 4300 rpm, 15 min, RT). Das Zellpellet wurde in 0,5 ml PBS gelöst und auf eine Blut-Agarplatte gegeben. Ebenfalls wurden als Kontrollen jeweils 10 ml einer üN-Kultur von dem Donor und dem Rezipienten einzeln pelletiert (Z! 4300 rpm, 15 min, RT), in 0,5 ml PBS gelöst und auf eine Blut-Agarplatte gegeben. Die Bakterienrasen wurden nach Inkubation üN bei 37°C von den Platten entfernt und in 1 ml PBS gelöst. Anschließend wurden Verdünnungsreihen erstellt, für

die Konjugation von 10^{-1} bis 10^{-3} und für die Kontrollen von 10^{-4} bis 10^{-6} . Es wurde jeweils das gesamte Volumen einer Verdünnungsstufe (450 µl) auf einer großen LB-Agarplatte (145 mm x 20 mm) ausplattiert. Die verwendeten Antibiotika für die Konjugation waren Chloramphenicol und Streptomycin bzw. Nalidixinsäure und für die Kontrollen entweder Chloramphenicol, Streptomycin oder Nalidixinsäure. Alle Agarplatten wurden üN inkubiert und die gewachsenen Kolonien gezählt und überprüft (HPI, phylogenetische Gruppe, β -Hämolysin).

Die Konjugationseffizienz wurde berechnet als Verhältnis von Transkonjuganten zu Donoren, in Kolonie-bildenden Einheiten pro Milliliter (cfu/ml) (Beaber et al., 2004).

$$\text{Konjugationseffizienz} = \frac{\text{Transkonjuganten}}{\text{Donoren}} \text{ cfu/ml}$$

9.3.2. Konjugation von aktiv transferierbaren Pathogenitätsinseln

Da die ICE-Typ-HPI Gene für den aktiven Transfer besitzt, wurde für die Konjugation keine Hilfe, wie etwa ein F'-Plasmid, benötigt. Wichtig war allerdings, die optimalen Konditionen für den Transfer zu ermitteln. Auch dieser Transfer konnte durch die Markierung der Insel mit einer Antibiotika-Resistenzkassette verfolgt werden. Der Versuch bestand aus mehreren Schritten.

a) Das ICE des zukünftigen Donors wurde mit einer Chloramphenicol- bzw. Kanamycin-Resistenzkassette nach der Methode von Datsenko und Wanner (Datsenko und Wanner, 2000) markiert (siehe "B.10.6. Insertion von Antibiotika-Resistenzkassetten"). Es wurden PCRs mit den Primern P1-fyuA-2104.for und P2-fyuA-2691.rev durchgeführt, um die Antibiotika-Resistenzkassetten zu amplifizieren. Für Chloramphenicol wurde als DNA-Vorlage das Plasmid pKD3, für Kanamycin das Plasmid pKD4 verwendet. Die PCR-Produkte wurden aufgereinigt und dialysiert. Es wurden aus üN-Kulturen der *E. coli*-WT-Stämme ED1a und EDF72 und des *K. pneumoniae*-WT-Stammes HU86 elektrokompetente Zellen hergestellt. In diese Stämme wurden je 2 µl des Plasmids pKD46 transformiert. Diese wurden anschließend üN bei 30°C und 200 rpm angezüchtet. Von diesen Kulturen wurden jeweils 1 ml in 40 ml frisches LB-Medium mit Arabinose (10 mM) und Ampicillin gegeben. Nach der Inkubation bei 30°C und 200 rpm wurden die Zellen bei einer OD₆₀₀ von ca. 0,5 geerntet. Nachdem die

Zellen elektrokompotent gemacht wurden, wurde in diese je 5 µl des PCR-Produkts transformiert. Die Stämme HU86 und ED1a erhielten eine Chloramphenicol-Resistenzkassette. Da der Stamm EDF72 bereits Chloramphenicol-resistent ist, erhielt dieser eine Kanamycin-Resistenzkassette.

b) Es wurden folgende Konditionen für einen erfolgreichen Transfer getestet:

1. Donoren, Rezipienten und Volumina: Als Donoren wurden die Stämme *K. pneumoniae* HU86 Cm, *E. coli* ED1a Cm und *E. coli* EDF72 Kn verwendet, als Rezipienten die Stämme *E. coli* MG1655 und *Y. enterocolitica* MRS40. Diese wurden in verschiedenen Kombinationen und in verschiedenen Volumina zwischen 1 ml und 50 ml für die Konjugation verwendet.
2. Wachstumsphasen: Die eingesetzten Bakterienstämme wurden in den exponentiellen und in den stationären Phasen untereinander kombiniert, d.h. Wachstumsphase Donor/Rezipient: stationär/stationär, stationär/exponentiell, exponentiell/stationär, exponentiell/exponentiell.
3. Temperaturen: Die getesteten Inkubationstemperaturen für die Konjugation waren 4°C, 30°C und 37°.
4. Dauer der Inkubation: Die Zeiten, in denen Donor und Rezipient zusammen inkubiert wurden, waren 0,5 h, 2 h, 3 h, 5 h, üN, 2 d und 5 d.
5. Nährmedien: Es wurde auf verschiedenen Festmedien (Blutagar, LB-Agar) und in Flüssigmedien (LB-Medium, mit/ohne schütteln, mit/ohne α, α' -Dipyridyl (100 mM; Sigma-Aldrich; St. Louis, MO, USA; Kat. # STBC4491V) konjugiert.
6. Konjugationsfilter: Die Konjugationen wurden mit und ohne Nitrocellulose Filter (MMF-Membrane Filters, Filter Typ 0,025 µm VSWP, Millipore (Merck; Darmstadt, Deutschland), Kat. # VSWP04700) durchgeführt.
7. SOS-Antwort im Donor: Der Donor wurde in eine Stresssituation versetzt, ausgelöst durch hohe Temperatur (50°C für 10 min, 30 min und 60 min), Zugabe von Nalidixinsäure (5 µg/ml, 10 µg/ml und 20 µg/ml), Zugabe von 1 N HCl (1%, 5%, und 10% v/v), osmotischer Druck (50%, 70%, 100% H₂O_{dest} v/v) und eine hohe Zelldichte.

- c) Nach der Konjugation wurde die Bakteriensuspension zur Selektion der Transkonjuganten auf Festmedium mit entsprechenden Antibiotika ausplattiert.
- d) Nach der Inkubation wurden die erhaltenen Transkonjuganten auf ihre Echtheit geprüft. Dafür wurde mittels MALDI die Gattung und Art des Bakteriums bestimmt. Das Vorhandensein der Antibiotika-Resistenz-markierten HPI wurde durch eine PCR bestimmt. Außerdem wurde mittels Phoenix die Antibiotikaresistenzen der Donoren, Rezipienten und Transkonjuganten ermittelt.
- e) Die bestätigten Transkonjuganten wurden auf die Funktionalität der ICE-Typ-HPI überprüft, da der Transfer über die Gattungsgrenze hinaus verlief. Dafür wurde der CAS-Agar zur Detektion von Siderophoren verwendet. Die Transkonjuganten wurden von einer üN bebrüteten Agarplatte auf eine CAS-Agarplatte ausgestrichen, welche anschließend üN kultiviert wurde.
- f) In welche tRNA das ICE integrierte, wurde für *E. coli* per PCR und für *Y. enterocolitica* nach Entschlüsselung des Genoms durch NGS bioinformatisch ermittelt.

10. Molekularbiologische und genetische Methoden

Diese Methoden beinhalten vor allem das Arbeiten mit DNA. Dabei geht es um deren Isolierung, Vervielfältigung und Manipulation durch Insertion von Antibiotika-Resistenzkassetten.

10.1. Isolierung von DNA für die PCR

Für PCRs wurden gekochte Zellen verwendet, da hierfür verhältnismäßig wenig DNA benötigt wird. Zur Herstellung wurde eine Kolonie von einer üN bebrüteten Agarplatte in ein Eppendorf-Reagenzgefäß überführt und in 100 µl H₂O_{dest} gelöst. Anschließend wurde die Lösung für 10 min bei 100°C gekocht um die Bakterienzellen aufzuschließen und die DNA zugänglich zu machen.

10.2. Isolierung von Plasmiden

Die extrachromosomale DNA, in diesem Fall Plasmid-DNA, wurde mit dem *QIAprep Spin Miniprep Kit* oder dem *Nucleo Bond® PC 20 Kit* laut den Angaben des Herstellers isoliert.

10.3. Polymerase-Kettenreaktion (PCR)

Für die spezifische Amplifikation von DNA-Abschnitten wurde eine Polymerase-Kettenreaktion (PCR) mit den dafür benötigten Primern durchgeführt. Dies funktioniert durch mehrere sich wiederholenden Schritte. Zuerst wird die DNA denaturiert (Denaturierung), dann binden die spezifischen Primer an die DNA (Hybridisierung) und anschließend fügt das Enzym DNA-Polymerase Nukleotide an die Primer-Enden (Elongation). Für die PCR wurde das *Taq DNA Core Kit* verwendet. Die Reagenzien und deren Volumen für eine einzelne PCR-Probe (25 µl) sind in Tabelle 11 aufgeführt.

Tabelle 11: Zusammensetzung einer einzelnen PCR mit 25µl Volumen.

Reagenzien (Konzentration)	Volumen
Forward-Primer (20 µM)	0,5 µl
Reverse-Primer (20 µM)	0,5 µl
10x Puffer ohne MgCl ₂	2,5 µl
MgCl ₂ Lösung (25 mM)	4 µl
10x dNTPs (jeweils 10 mM)	1 µl
DNA (oder Bakterienkolonie)	1,25 – 12,5 µg
<i>Taq</i> Polymerase	0,2 µl
H ₂ O _{dest}	ad 25 µl

Für die Amplifikation der Antibiotika-Resistenzkassetten von den Plasmiden pKD3 (Chloramphenicol) und pKD4 (Kanamycin) wurde zusätzlich die *Pfu* Polymerase (Thermo Fisher Scientific; Schwerte, Deutschland; Kat. # EP0572) verwendet, da diese eine Fehlerkorrekturfunktion (*Proof-Reading*) besitzt. In diesem Fall betrug das Volumen einer einzelnen PCR 50 µl, d.h. alle Reagenzien wurden mit doppeltem Volumen eingesetzt und zu den 0,2 µl *Taq* Polymerase kamen zusätzliche 0,2 µl *Pfu* Polymerase hinzu.

Die Amplifikation erfolgte in einem PCR-Thermocycler. Es wurden Primer-spezifische Hybridisierungstemperaturen verwendet und die Dauer der Elongation wurde je nach Größe

des zu amplifizierenden DNA-Abschnittes gewählt. Der Ablauf einer PCR ist in Tabelle 12 aufgeführt.

Tabelle 12: Ablauf einer PCR.

Schritt	Zeit und Temperatur
1) Initiale Denaturierung	10 min, 95°C
2) Denaturierung	45 s, 95°C
3) Hybridisierung	45 s, 50 - 60°C
4) Elongation	1 min / kb, 72°C
5) Finale Elongation	10 min, 72°C
6) Abkühlung	∞, 4°C
Schritte 2 bis 4: 35 Wiederholungen	

10.4. Trennung, Aufreinigung und Konzentrationsbestimmung von Nukleinsäuren

Die Auftrennung von DNA-Fragmenten, die z.B. durch eine PCR oder einen enzymatischen Verdau entstanden, erfolgte mittels horizontaler Gelelektrophorese in Flachbett-Elektrophorese-Kammern mit einem 1,5%igen Agarosegel. Die Gele entstanden durch die Zugabe von 1,5% (w/v) Agarose (Carl-Roth; Karlsruhe, Deutschland; Kat. # 2267.4) in TAE-Puffer mit anschließender Aufschmelzung in der Mikrowelle. Nachdem das flüssige Gel auf ca. 50°C abgekühlt wurde, wurde diesem 3% (v/v) Ethidiumbromid hinzugegeben, es wurde durchmischt und zum Erkalten in eine passende Form mit einem entsprechenden Kamm gegeben. Das fertige Gel wurde in einer Elektrophorese-Kammer mit TAE-Puffer bedeckt und der Kamm entfernt. Die zu untersuchenden Proben wurden mit 6x DNA Auftragspuffer (Loading Dye™; Thermo Fisher Scientific; Schwerte, Deutschland; Kat. # R0611) vermischt und in die durch den Kamm entstandenen Geltaschen gefüllt. Zur späteren Identifizierung der DNA-Fragmente wurde zusätzlich ein Größenmarker (GeneRuler™ 100 bp oder GeneRuler™ 1 kb; Thermo Fisher Scientific; Schwerte, Deutschland; Kat. # SM0321 und SM0311) aufgetragen. Die Gelelektrophorese lief bei 120 V für 45 min bei PCR-Produkten und bei 80 V für 90 min bei Kontrollverdauen. Die aufgetrennte DNA wurde unter UV-Licht sichtbar gemacht und digital dokumentiert.

Positive PCR-Produkte wurden bei Bedarf aufgereinigt. Hierfür wurde das *QIAquick PCR Purification Kit* nach den Angaben des Herstellers verwendet. Wenn nötig, wurden die zu

untersuchenden Banden eines Gels ausgeschnitten und mit dem *Ultra Clean Gel Spin – DNA Extraction Kit* aufgereinigt.

Zur Konzentrations- und Qualitätsbestimmung der DNA wurde das NanoDrop 2000 Spektralphotometer verwendet. Mit einer Wellenlänge von 260 nm, dem Absorptionsmaximum von Nukleinsäuren, konnte die Konzentration von DNA bestimmt werden. Dabei entspricht eine OD₂₆₀ von 1 einer DNA-Konzentration von 50 µg/ml. Eine Wellenlänge von 280 nm gibt die Konzentration von Proteinen an, eine von 230 nm die von Polysacchariden. Um die Reinheit der DNA zu überprüfen, wurden auch die Verhältnisse von 260/280 nm und von 260/230 nm verglichen. Der Reinheitsgrad der DNA ist bei Werten zwischen 1,8 – 2,0 (260/280 nm) und 1,8 - 2,2 (260/230 nm) am höchsten.

10.5. Verdau von Nukleinsäuren mit Restriktionsendonukleasen

Um DNA zu verdauen wurden Restriktionsendonukleasen mit ihren entsprechenden Puffern verwendet. Es wurde pro Ansatz 30 µl Gesamtvolumen eingesetzt, davon 10% (v/v) spezifische Pufferlösung, in H₂O_{dest} gelöste DNA und 1 Unit (Einheit, U) Enzym pro µg DNA. Der Verdau fand üN bei der für das Enzym optimalen Temperatur unter leichtem Schütteln von 300 rpm in einem Schüttelinkubator statt. Manche Restriktionsenzyme mussten nach dem Verdau bei 65°C für 20 min hitzeinaktiviert werden. Der Restriktionsansatz wurde anschließend mittels Gelelektrophorese überprüft und dokumentiert.

10.6. Insertion von Antibiotika-Resistenzkassetten

Die Insertion einer Antibiotika-Resistenzkassette erfolgte nach dem Prinzip der von Kirill Datsenko und Barry Wanner beschriebenen Methode (Datsenko und Wanner, 2000), in der eine chromosomale Region durch eine beliebige DNA Sequenz durch homologe Rekombination ausgetauscht werden kann. Diese Rekombination kommt durch die Red Rekombinase des λ Phagen zustande. Um die HPI zu markieren, wurde als Zielregion eine nicht-codierende Stelle *downstream* von *fyuA* gewählt, um den Einfluss auf die nachfolgenden Transfers zu minimieren. Hierfür wurde erst ein Primer-Paar generiert, das sowohl eine zu der chromosomalen Zielregion als auch zu der Resistenz Kassette auf dem Plasmid pKD3 (Chloramphenicol) bzw. pKD4 (Kanamycin) homologe Sequenz enthält. Eine PCR auf ein entsprechendes Plasmid führte zu einem PCR-Produkt, das nach der Transformation mit Hilfe des bereits im Vorfeld transformierten Plasmids pKD46 in der Lage war, eine chromosomale

Zielregion zu ersetzen. Dies wurde ermöglicht durch die Zugabe von Arabinose (10 mM) in das Flüssigmedium, welches die bereits erwähnte Red Rekombinase auf dem Plasmid pKD46 induziert. Anschließend wurden die Bakterien auf eine erfolgreiche Integration der Resistenzkassette auf einer LB-Platte mit dem entsprechenden Antibiotikum selektiert und anschließend per PCR und Sequenzierung auf Richtigkeit überprüft.

10.7. Sequenzierung von kleinen DNA Abschnitten

Um einerseits die Richtigkeit der Insertion einer Antibiotika-Resistenzkassette zu überprüfen, und andererseits kleine Lücken innerhalb des *whole genome sequencing* zu schließen, wurden kleine DNA Abschnitte sequenziert. Die Sequenzierungen wurden von der Firma Eurofins (Ebersberg, Deutschland) durchgeführt. Hierfür wurde ein aufgereinigtes PCR-Produkt (ca. 10 ng/µl) in einem Volumen von 15 µl mit den passenden Sequenzierungs-Primern (10 µl, 5 µM) eingesendet. Die fertigen Sequenzen wurden mit der Software CLC Genomics Workbench ausgewertet.

11. Next-Generation Sequencing (NGS)

In den letzten Jahren konnten durch *Next-Generation Sequencing* (NGS) gesamte Genomsequenzen von vielen Lebewesen bestimmt werden. Ein großer Teil dieser Arbeit beschäftigte sich mit dieser Methode und der Auswertung der entstandenen Rohdaten.

11.1. Auswahl und Anzucht der Bakterien

Für den phylogenetischen Abschnitt der drei immobilen PAIs (HPI, *pks*- und *serU*-Insel) wurde größtenteils die ECOR-Sammlung zum *whole genome sequencing* ausgewählt, aber auch die ebenfalls bereits publizierten *E. coli*-Stämme S107 und S108. Für die sowohl mobilen als auch immobilen PAI-Transfers wurden jeweils die Donoren, Rezipienten und einige Transkonjuganten für die Sequenzierung ausgewählt. Alle Bakterien wurden in Flüssigmedium angezüchtet, um die Zellen in der stationären Wachstumsphase ernten zu können. In dieser Phase ist die Teilung eingestellt und das Chromosom gleichmäßig verteilt. Somit sollte eine gleichmäßige Sequenzierung des Genoms erreicht werden.

11.2. Isolierung und Qualitätsüberprüfung der genomischen DNA (gDNA)

Für die Sequenzierung der Bakterienstämme musste erst deren gDNA gewonnen werden. Hierfür wurde 1 ml einer üN-Kultur bei 5.000 rpm für 5 min zentrifugiert und der Überstand verworfen. Zur Isolierung der gDNA wurde das *High Pure PCR Template Preparation Kit* nach Angaben des Herstellers verwendet. Die DNA wurde in 100 µl H₂O_{dest} eluiert und mit dem NanoDrop 2000 Spektralphotometer vermessen. Nur diejenige gDNA, die dem Reinheitsgrad entsprach, wurde weiterverwendet.

11.3. Überprüfung auf Verdaubarkeit der genomischen DNA

Die gDNA, die zur Sequenzierung ausgewählt wurde, musste erst auf Verdaubarkeit überprüft werden. Hierfür wurde 1 µg gDNA mit 1 U *EcoRI* und *EcoRI*-Puffer üN bei 37°C verdaut. Am nächsten Tag wurde der Verdau auf ein Gel aufgetragen und mittels Gelelektrophorese überprüft. Wenn die DNA komplett verdaut wurde, war ein kontinuierlich über viele Größenbereiche verlaufende DNA-Bande (Schmier) auf dem Gel zu sehen (Abb. 11).

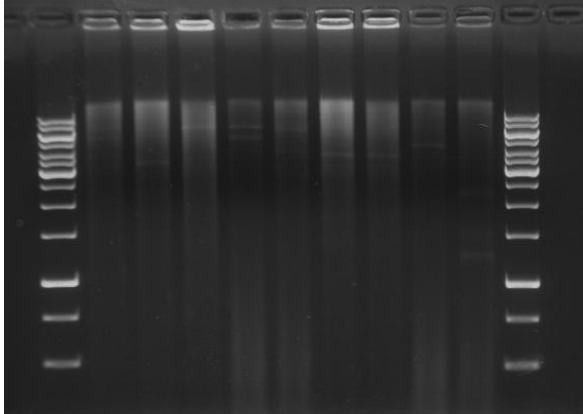


Abbildung 11: Gel-Bild eines üN-Verdau von gDNA mit *EcoRI*.

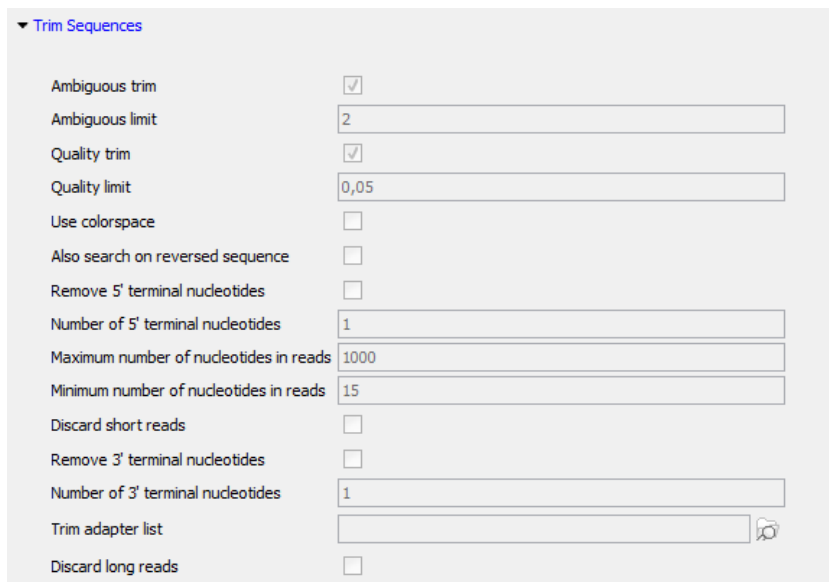
11.4. Sequenzierung

Etwa 20 µg reine und verdaubare gDNA wurde in einem Volumen von 100 µl (H₂O_{dest}) für die Sequenzierung an den Kooperationspartner Philippe Glaser (Institut Pasteur; Paris, Frankreich) gesendet. In seiner Arbeitsgruppe wurde das *TruSeq Kit* (Illumina; San Diego, CA, USA) verwendet um die sogenannten *Libraries* zu erstellen. Der Sequenzierer, der benutzt wurde, war ein HiSeq 2000 (Illumina; San Diego, CA, USA). Die maximale *Read*-

Länge betrug 100 Nukleotide (nt) *single-read*. Die Rohdaten wurden zur weiteren Bearbeitung zugesandt.

11.5. Verwendung der Software CLC Genomics Workbench

Für alle bioinformatischen Auswertungen wurde die Software CLC Genomics Workbench 6.5 verwendet. Die Rohdaten wurden als Illumina-Daten in das Programm importiert (Parameter: "*Remove failed reads*", "*Quality scores: NCBI/Sanger or Illumina Pipeline 1.8 and later*"). Um die Rohdaten zu assemblieren wurden zwei *Workflows* generiert: "Assemblierung an Referenz" und "*de novo* Assemblierung". Die Parameter für das "Trimmen" waren bei beiden *Workflows* gleich und sind in Abbildung 12 zu sehen.




▼ Trim Sequences	
Ambiguous trim	<input checked="" type="checkbox"/>
Ambiguous limit	2
Quality trim	<input checked="" type="checkbox"/>
Quality limit	0,05
Use colorspace	<input type="checkbox"/>
Also search on reversed sequence	<input type="checkbox"/>
Remove 5' terminal nucleotides	<input type="checkbox"/>
Number of 5' terminal nucleotides	1
Maximum number of nucleotides in reads	1000
Minimum number of nucleotides in reads	15
Discard short reads	<input type="checkbox"/>
Remove 3' terminal nucleotides	<input type="checkbox"/>
Number of 3' terminal nucleotides	1
Trim adapter list	<input type="text"/> 
Discard long reads	<input type="checkbox"/>

Abbildung 12: Verwendete Parameter für das Trimmen der *Reads* innerhalb der *Workflows*.

Assemblierung an Referenz: Hierbei wurden die *Reads* an ein Referenz-Genom oder -Plasmid angelegt. Als bakterielles Genom wurde meist die Sequenz *E. coli*-Stammes 536 (NCBI-Zugangsnummer NC_008253.1) gewählt, weil diese alle zu untersuchenden Inseln beinhaltet. Dieser *Workflow* wurde v.a. bei den phylogenetischen Auswertungen verwendet. Falls kein entsprechender *Read* an die Referenz angelegt werden konnte, wurde stattdessen ein "N" eingefügt. In Abbildung 13 sind die verwendeten Parameter zu sehen.

▼ All parameters

Masking mode	No masking
Masking track	
Mismatch cost	2
Insertion cost	3
Deletion cost	3
Length fraction	0,5
Similarity fraction	0,8
Global alignment	<input type="checkbox"/>
Color space alignment	<input checked="" type="checkbox"/>
Color error cost	3
Auto-detect paired distances	<input checked="" type="checkbox"/>
Non-specific match handling	Map randomly

Abbildung 13: Verwendete Parameter für den *Workflow* "Assemblierung an Referenz".

de novo Assemblierung: In diesem *Workflow* wurden die 100 nt *Reads* als längere *Contigs* zusammengefügt. Dies wurde v.a. für die Transkonjuganten Auswertungen angewendet. Die verwendeten Parameter sind in Abbildung 14 zu sehen.

▼ De Novo Assembly

Mapping mode	Map reads back to contigs (slow)
Update contigs	<input checked="" type="checkbox"/>
Automatic bubble size	<input checked="" type="checkbox"/>
Bubble size	50
Minimum contig length	200
Automatic word size	<input checked="" type="checkbox"/>
Word size	20
Create report	<input type="checkbox"/>
Guidance only reads	
Perform scaffolding	<input checked="" type="checkbox"/>
Auto-detect paired distances	<input checked="" type="checkbox"/>
Mismatch cost	2
Insertion cost	3
Deletion cost	3
Length fraction	0,5
Similarity fraction	0,8
Alignment mode	local
Colourspace alignment	<input checked="" type="checkbox"/>
Colourspace error cost	3
Create list of un-mapped reads	<input type="checkbox"/>

Abbildung 14: Verwendete Parameter für den *Workflow* "de novo Assemblierung".

Mit den erstellten Sequenzen wurden nun Stammbäume oder andere Sequenz-Vergleiche erstellt. Im Folgenden sind die verschiedenen Methoden und die verwendeten Parameter kurz zusammengefasst.

Alignment: Um Sequenzen miteinander zu vergleichen wurde ein *Alignment* erstellt. Es wurden mindestens zwei Sequenzen hierfür benötigt. Die verwendeten Parameter waren "Gap open cost: 10,0", "Gap extension cost: 1,0" und "Less accurate".

Stammbäume: Zur Erstellung der Stammbäume wurde ein *Alignment* ausgewählt und mit folgenden Parametern zusammengefügt: "Neighbor Joining", "Jukes-Cantor", "Perform bootstrap analysis" und "Replicates: 100".

Statistik: Um statistische Auswertungen durchzuführen wurde die Funktion "Pairwise Comparison" verwendet. Es wurden die Berechnungen der "Differences" (Anzahl an SNPs) und "Percent identity" (prozentuale Homologie) ausgewertet.

DNA-Zugehörigkeit innerhalb der Transkonjuganten: Zur Unterscheidung zwischen Donor- und Rezipienten-DNA innerhalb der Transkonjuganten wurde ein *Alignment* der Donor-, Rezipienten- und Transkonjuganten-Sequenz erstellt und verglichen. Dabei wurde die Stelle der Rekombination als Bereich zwischen der letzten bekannten „Donor-Base“ und der letzten Base vor beginnender "Rezipienten-DNA" definiert (Abb. 15). Die NCBI-Zugangsnummer für die Sequenz des Stammes MG1655 war U00096.

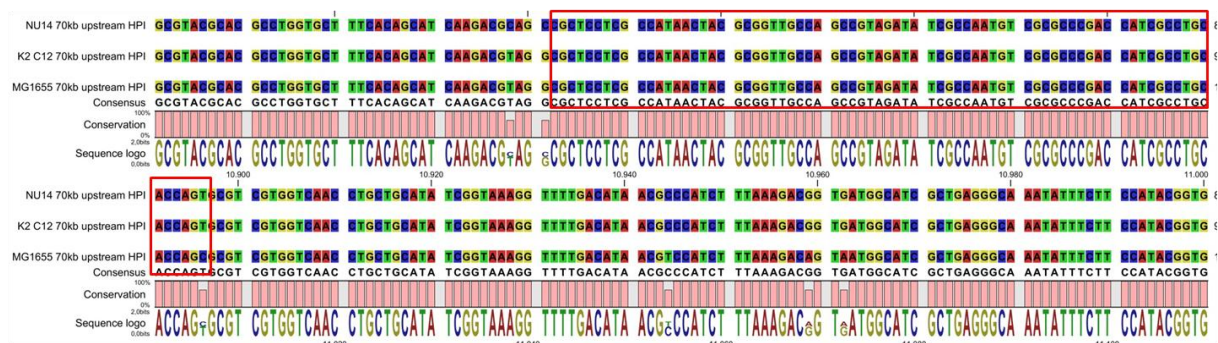


Abbildung 15: Beispiel für die DNA-Zugehörigkeit innerhalb der Transkonjuganten *upstream* der HPI. Von oben nach unten zu sehen sind der Donor (NU14), der Transkonjugant (K2) und der Rezipient (MG1655). Der rote Rahmen zeigt den Bereich der Rekombination. Dieser ist definiert als Bereich zwischen der letzten bekannten Base des Donors (links unten) und der Base, die vor der ersten bekannten Base des Rezipienten liegt (oben Mitte).

F'-Plasmid der Transkonjuganten: Um zu überprüfen ob der zu untersuchende Transkonjugant noch ein F'-Plasmid trägt, wurden die *Reads* an die F'-Plasmid Referenz (NCBI-Zugangsnummer AP001918) angelegt. Wenn die *Reads* vollständig zugeordnet werden konnten, war das Plasmid noch in dem Transkonjuganten vorhanden.

C. Ergebnisse

Die Familie der *Enterobacteriaceae* besitzt eine Vielzahl an Pathogenitätsinseln (PAIs), die für unterschiedliche Krankheitsbilder beim Menschen verantwortlich sind. Diese PAIs können an apathogene Bakterienstämme weitergegeben werden und sich somit verbreiten (Dobrindt et al., 2004). Der Mechanismus dieser Übertragung wurde bereits intensiv untersucht, konnte aber noch nicht eindeutig geklärt werden (Johnson und Grossman, 2015; Schubert et al., 2009).

Das Ziel dieser Arbeit war es, weitere Einblicke in die Transfermechanismen von PAIs zu bekommen. Einerseits wurden Inseln untersucht, die durch das Vorhandensein gewisser Gene den Donor zu einem aktiven Transfer befähigen. Diese PAIs werden auch integrative und konjugative Elemente (*integrative and conjugative elements*, ICEs) genannt (Schubert et al., 2004a). Andererseits wurde analysiert, wie Inseln ohne diese Gene verbreitet werden können. Im Mittelpunkt der Untersuchungen stand die relativ junge Methode *Next-Generation Sequencing* (NGS).

Durch die Sequenzierung von ganzen Genomen wurde die gesamte *Escherichia coli* Referenz (ECOR)-Sammlung *in silico* untersucht. Mit dieser Sammlung wurde sowohl der Transfer von immobilen PAIs als auch die Phylogenie der Art *E. coli* analysiert. Einige ICE-tragende Isolate wurden ebenfalls für die Genom-Entschlüsselung durch NGS ausgewählt. Deren PAIs wurden *in silico* miteinander verglichen, um die Verwandtschaft der ICEs zu bestimmen.

Die Übertragung der unterschiedlichen PAIs wurde ebenfalls *in vitro* durch Konjugationsversuche nachgestellt und analysiert. Für den Transfer der immobilen Inseln wurde die Hilfe eines F'-Plasmids benötigt, welches zusätzlich das benachbarte *Backbone*-Genom mit überträgt (Frost et al., 1994). Für die Übertragung der ICEs mussten erst geeignete Konjugationsbedingungen untersucht werden, weil sich dieser Transfer als ineffizient erwies (Johnson und Grossman, 2015). Zuletzt wurden die erhaltenen Transkonjuganten ebenfalls sequenziert und *in silico* ausgewertet.

1. Next-Generation Sequencing

Im Rahmen dieser Arbeit wurden insgesamt 100 bakterielle Genome mittels NGS entschlüsselt. Zum ersten Mal wurde die gesamte ECOR-Sammlung mit 72 Stämmen komplett sequenziert. Dies wurde mit einem Illumina HiSeq 2000 bei einer maximalen *Read*-Länge von 100 nt *single-read* durchgeführt. Im Durchschnitt bestanden die Rohdaten aus 9.105.077 *Reads* und in 99,84% konnten keine mehrdeutigen Basen gefunden werden. Der Phred-Qualitätswert betrug durchschnittlich 40. Dieser Wert sagt aus, dass die Wahrscheinlichkeit einer falsch eingebauten Base bei 0,001% liegt. Die *de novo* Assemblierung erbrachte im Durchschnitt 150 *Contigs*, die größer als 1 kb waren, mit maximalen Längen von 145 - 430 kb. Die PAIs *High-Pathogenicity Island* (HPI), *pks*- und *serU*-Insel waren in allen Genom-Sequenzen komplett vorhanden und lagen in *Contigs* mit teilweise 400 kb vor. Bei den *Housekeeping*-Genen *dinB* und *uidA*, die für die Untersuchung der *E. coli*-Phylogenie genutzt wurden, traten kleinere Lücken auf, sodass jeweils nur ein bestimmtes Fragment dieser Gene verwendet werden konnte. In dem Bereich der PAI III fehlten große Sequenz-Abschnitte, was womöglich durch die Überreste von IS-Elementen und Transposons innerhalb dieser Insel zustande gekommen sein könnte (Dobrindt et al., 2002).

2. Ergebnisse der durchgeführten PCRs und Einteilung in PAI-Gruppen

Alle verwendeten Stämme wurden vor der Genom-Sequenzierung auf die Anwesenheit der PAIs HPI, PAI III, *pks*- und *serU*-Insel getestet. Bei den *E. coli*-Stämmen wurden zusätzlich die phylogenetischen Gruppen ermittelt. Die durchgeführten Polymerase-Kettenreaktionen (PCRs) dienten zum späteren Vergleich mit den *in silico*-Daten der Sequenzierung. Für die Bestimmung der phylogenetischen Gruppe wurde die Triplex-PCR nach Clermont *et al.* mit den Primern ChuA.1, ChuA.2, YjaA.1, YjaA.2, TspE4C2.1 und TspE4C2.2 durchgeführt (Clermont et al., 2000). Die HPI wurde mit den Primern *fyuA.1080.for* und *fyuA.1709.rev* nachgewiesen. Für die *pks*-Insel wurden die Primer *clbA.IHE.for* und *clbA.IHE.rev* und für die *serU*-Insel die Primer *tcpC.for* und *tcpC.rev* eingesetzt. Der Nachweis der PAI III erfolgte mit den Primern *iroN.28.for* und *iroN.2053.rev*. In Tabelle 13 sind die Ergebnisse für die 75 untersuchten Stämme, die für die Phylogenie der immobilen PAIs verwendet wurden, aufgeführt.

Tabelle 13: PCR-Ergebnisse der 75 *E. coli*-Stämme zur phylogenetischen Untersuchung der immobilisierenden PAIs. Aufgelistet sind die phylogenetischen Gruppen und die PAIs HPI, *pks*-Insel, *serU*-Insel und PAI III.

Stamm	phyl. Gr.	HPI	<i>pks</i> -Insel	<i>serU</i> -Insel	PAI III
ECOR01	A	negativ	negativ	negativ	negativ
ECOR02	A	positiv	negativ	negativ	negativ
ECOR03	A	negativ	negativ	negativ	negativ
ECOR04	A	negativ	negativ	negativ	negativ
ECOR05	A	negativ	negativ	negativ	negativ
ECOR06	A	negativ	negativ	negativ	negativ
ECOR07	A	negativ	negativ	negativ	negativ
ECOR08	A	positiv	negativ	negativ	negativ
ECOR09	A	positiv	negativ	negativ	negativ
ECOR10	A	positiv	negativ	negativ	negativ
ECOR11	A	positiv	negativ	negativ	negativ
ECOR12	A	negativ	negativ	negativ	negativ
ECOR13	A	negativ	negativ	negativ	negativ
ECOR14	A	negativ	negativ	negativ	negativ
ECOR15	A	negativ	negativ	negativ	negativ
ECOR16	A	positiv	negativ	negativ	negativ
ECOR17	A	negativ	negativ	negativ	negativ
ECOR18	A	negativ	negativ	negativ	negativ
ECOR19	A	negativ	negativ	negativ	negativ
ECOR20	A	negativ	negativ	negativ	negativ
ECOR21	A	negativ	negativ	negativ	negativ
ECOR22	A	negativ	negativ	negativ	negativ
ECOR23	A	negativ	negativ	negativ	negativ
ECOR24	A	positiv	negativ	negativ	negativ
ECOR25	A	negativ	negativ	negativ	negativ
ECOR26	B1	negativ	negativ	negativ	negativ
ECOR27	B1	negativ	negativ	negativ	negativ
ECOR28	B1	negativ	negativ	negativ	negativ
ECOR29	B1	negativ	negativ	negativ	negativ
ECOR30	B1	negativ	negativ	negativ	negativ
ECOR31	D	positiv	negativ	negativ	negativ
ECOR32	B1	negativ	negativ	negativ	negativ
ECOR33	B1	negativ	negativ	negativ	negativ
ECOR34	B1	negativ	negativ	negativ	negativ
ECOR35	D	positiv	negativ	negativ	negativ
ECOR36	D	positiv	negativ	negativ	negativ
ECOR37	D	negativ	negativ	negativ	negativ
ECOR38	D	positiv	negativ	negativ	negativ
ECOR39	D	positiv	negativ	negativ	negativ
ECOR40	D	positiv	negativ	negativ	negativ
ECOR41	D	positiv	negativ	negativ	negativ
ECOR42	D	negativ	negativ	negativ	negativ
ECOR43	A	positiv	negativ	negativ	negativ

ECOR44	D	negativ	negativ	negativ	negativ
ECOR45	B1	negativ	negativ	negativ	negativ
ECOR46	D	positiv	negativ	negativ	negativ
ECOR47	D	negativ	negativ	negativ	negativ
ECOR48	D	positiv	negativ	negativ	negativ
ECOR49	D	positiv	negativ	negativ	negativ
ECOR50	D	positiv	negativ	negativ	negativ
ECOR51	B2	positiv	positiv	negativ	positiv
ECOR52	B2	positiv	positiv	negativ	positiv
ECOR53	B2	positiv	positiv	positiv	positiv
ECOR54	B2	positiv	positiv	negativ	positiv
ECOR55	B2	positiv	positiv	negativ	negativ
ECOR56	B2	positiv	positiv	positiv	negativ
ECOR57	B2	positiv	positiv	positiv	negativ
ECOR58	B1	negativ	negativ	negativ	negativ
ECOR59	B2	positiv	negativ	negativ	negativ
ECOR60	B2	positiv	positiv	positiv	positiv
ECOR61	B2	positiv	negativ	negativ	negativ
ECOR62	B2	positiv	negativ	negativ	positiv
ECOR63	B2	positiv	positiv	positiv	positiv
ECOR64	B2	positiv	negativ	negativ	negativ
ECOR65	B2	positiv	positiv	negativ	positiv
ECOR66	B2	positiv	negativ	negativ	positiv
ECOR67	B1	negativ	negativ	negativ	negativ
ECOR68	B1	positiv	negativ	negativ	negativ
ECOR69	B1	negativ	negativ	negativ	negativ
ECOR70	B1	positiv	negativ	negativ	positiv
ECOR71	B1	positiv	negativ	negativ	negativ
ECOR72	B1	positiv	negativ	negativ	negativ
536	B2	positiv	positiv	positiv	positiv
S107	B2	positiv	negativ	positiv	negativ
S108	B2	positiv	negativ	positiv	negativ

Auffällig ist, dass die *pks*-Insel und die *serU*-Insel nur zusammen mit der HPI und nur in B2-Stämmen vorhanden sind, was bereits in vorherigen Arbeiten berichtet wurde (Putze et al., 2009; Schubert et al., 2010). Alle B2-Stämme sind HPI-positiv. Die prozentuale Verteilung der PAIs und der PAI-Kombinationen ist in Tabelle 14 aufgelistet. Die Stämme wurden auch auf das Vorhandensein der Region X überprüft. Diese Region ist bei einigen ICE-Typ-HPIs zwischen der konservierten Kernregion und der *asn* tRNA zu finden und ist homolog zu dem Virulenzplasmid pLVPK von *Klebsiella pneumoniae* (Preisach, 2008). Der einzige Region X-positive Stamm ist ECOR31. Dies wurde mit den Primern RegX-521.for und RegX-358.rev nachgewiesen.

Tabelle 14: Auflistung der prozentualen PAI-Verteilung der 75 in Tabelle 13 aufgeführten *E. coli*-Stämme. Zuerst wurde die Verteilung der einzelnen PAIs, unabhängig von der HPI, aufgelistet. Im Weiteren wird die Verteilung der PAIs als Kombinationen untereinander beschrieben.

PAIs	prozentuale Verteilung (Anzahl Stämme)
keine PAI	45% (34)
insgesamt HPI	55% (41)
insgesamt <i>pks</i> -Insel	15% (11)
insgesamt <i>serU</i> -Insel	11% (8)
nur HPI	37% (28)
Kombination HPI und <i>pks</i> -Insel	7% (5)
Kombination HPI und <i>serU</i> -Insel	3% (2)
Kombination HPI, <i>pks</i> - und <i>serU</i> -Insel	8% (6)

Die PAIs *pks*- und *serU*-Insel sind nur in Verbindung mit der HPI und nur in der phylogenetischen Gruppe B2 präsent (Tab. 13). Da diese drei PAIs im *E. coli*-Genom zudem in unmittelbarer Nähe zueinander zu finden sind, ließ dies den Schluss zu, dass womöglich ein *en bloc*-Transfer dieser Inseln stattgefunden hatte und zu dieser Verteilung führte (Schubert et al., 2009; Putze et al., 2009; Cirl et al., 2008). Um dies zu analysieren, wurden diese PAIs am ausführlichsten untersucht. Die Stämme mit diesen Inseln wurden in sogenannte PAI-Gruppen eingeteilt. Stämme, die nur die HPI besitzen, wurden als PAI-Gruppe 1 definiert. Bakterien mit einer Kombination aus HPI und *pks*-Insel wurden als PAI-Gruppe 2a, und solche mit einer Kombination aus HPI und *serU*-Insel als PAI-Gruppe 2b bezeichnet. Bakterienstämme, die alle drei Inseln in ihrem Genom tragen, wurden als PAI-Gruppe 3 definiert.

Ebenfalls ist die PAI III nur in Anwesenheit der HPI zu finden, allerdings unabhängig von den anderen untersuchten PAIs und sowohl in B2- als auch B1-Stämmen. Die Verteilung ist in Tabelle 15 zu sehen. Aufgrund der Entfernung zu den anderen PAIs wurde davon ausgegangen, dass die PAI III nicht *en bloc* mit diesen Inseln übertragen wurde.

Tabelle 15: Verteilung der PAI III innerhalb der 75 untersuchten Stämme, die für die phylogenetischen Auswertungen der immobilisierten PAIs verwendet wurden.

PAIs	prozentuale Verteilung (Anzahl Stämme)
insgesamt PAI III	15% (11)
Kombination HPI und PAI III	4% (3)
Kombination HPI, <i>pks</i> -Insel und PAI III	5% (4)
Kombination HPI, <i>serU</i> -Insel und PAI III	0% (0)
Kombination HPI, <i>pks</i> -, <i>serU</i> -Insel und PAI III	5% (4)

Die *E. coli*-Stämme, die für den Transfer der immobilen PAIs verwendet wurden, wurden ebenfalls per PCR auf die Anwesenheit der PAIs überprüft. Der Donor-Stamm 536 ist bereits in Tabelle 13 zu sehen. Der zweite Donor-Stamm, NU14, gehört zur phylogenetischen Gruppe B2 und besitzt die HPI, *pks*-Insel und PAI III. Die Rezipienten-Stämme, MG1655 und MG1655 Nal, sind in der phylogenetischen Gruppe A und besitzen keine PAI.

Die Stämme der Familie *Enterobacteriaceae*, die für den Transfer der mobilen PAI (ICE) genutzt wurden, wurden ebenfalls per PCR untersucht. Für den Nachweis der HPI wurden die Primer *fyuA*.1080.for und *fyuA*.1709.rev und für den Nachweis der *pks*-Insel die Primer *clbA*.IHE.for und *clbA*.IHE.rev verwendet. Die phylogenetische Gruppe der *E. coli*-Stämme wurde mit den Primern *ChuA*.1, *ChuA*.2, *YjaA*.1, *YjaA*.2, *TspE4C2*.1 und *TspE4C2*.2 bestimmt (Clermont et al., 2000). Die Region X wurde mit den Primern *RegX*-521.for und *RegX*-358.rev nachgewiesen. Der Donor-Stamm *K. pneumoniae* HU86 hat eine ICE-Typ-HPI mit Region X. Die *E. coli*-Donor-Stämme ED1a und EDF72 sind in der phylogenetischen Gruppe B2 und besitzen eine ICE-Typ-HPI ohne Region X. Die Rezipienten-Stämme sind alle HPI-negativ und bezüglich *E. coli* in der phylogenetischen Gruppe A.

3. Phylogenetische Auswertungen der Housekeeping-Gene und der immobilen PAIs

Zur besseren Hervorhebung innerhalb der Stammbäume wurden die vier PAI-Gruppen in einen entsprechenden Farb-Code eingeteilt. PAI-Gruppe 1 (HPI) wurde die Farbe Lila zugewiesen, PAI-Gruppe 2a (HPI, *pks*-Insel) die Farbe Blau, PAI-Gruppe 2b (HPI, *serU*-Insel) Grün und die PAI-Gruppe 3 (HPI, *pks*-Insel, *serU*-Insel) wurde in Rot dargestellt. Für die vier phylogenetischen Gruppen wurden verschiedene Symbole verwendet. Die Gruppen B2 und D, denen eine höhere Pathogenität innerhalb der extraintestinal pathogenen *E. coli* (ExPEC) zugesprochen wird, wurden als Punkt bzw. als Kreis dargestellt. Die Gruppen A und B1, die als eher weniger pathogen innerhalb der ExPEC gelten, wurden als Quadrat und als Raute abgebildet (Clermont et al., 2000). Somit kann man auf einen Blick erkennen, welcher PAI-Gruppe und welcher phylogenetischen Gruppe ein Bakterienstamm angehört. Der Farb-Code und die Symbole sind in Abbildung 16 zu sehen.

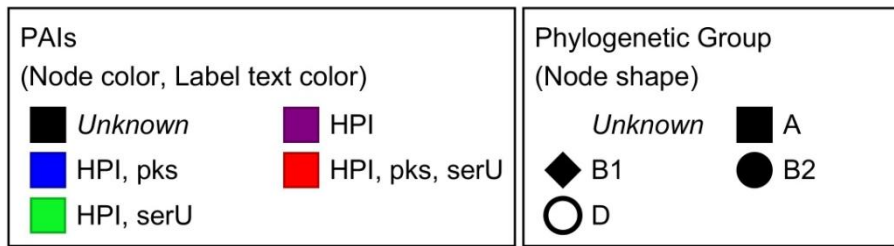


Abbildung 16: Farb-Code der unterschiedlichen PAI-Gruppen und Symbole der phylogenetischen Gruppen.

Mit einer Multilocus-Sequenz-Typisierung (MLST) werden unterschiedliche Abschnitte eines bakteriellen Genoms mit denen anderer Genome verglichen. Damit kann die Homologie und der Verwandtschaftsgrad dieser Abschnitte bestimmt werden. Zur Untersuchung der Phylogenie existieren verschiedene MLST-Schemata, welche unterschiedliche *Housekeeping*-Genfragmente einzelner Bakterienstämme miteinander vergleichen (Wirth et al., 2006; Jaureguy et al., 2008). Für die phylogenetische Untersuchung wurden sechs *Housekeeping*-Gene sowie deren Fragmente bzw. die Fragmente von acht *Housekeeping*-Genen untersucht. Von den drei PAIs HPI, *pks*- und *serU*-Insel wurde jeweils eine MLST, bestehend aus sechs codierenden PAI-Genen, angefertigt und mit den Stammbäumen der ganzen Inseln verglichen. Zusätzlich wurde eine statistische Auswertung bezüglich Homologie und Einzelnukleotid-Polymorphismus (SNP) aller Sequenzen durchgeführt.

3.1. Housekeeping-Gene

Anhand der *Housekeeping*-Gene können Bakterien in phylogenetische Gruppen eingeteilt werden. Es gibt verschiedene MLST-Schemata, in denen unterschiedliche *Housekeeping*-Genfragmente verglichen werden (Jaureguy et al., 2008; Wirth et al., 2006). Das MLST-Schema, welches verwendet wurde, wurde von einer Arbeitsgruppe des Pasteur-Instituts (Paris, Frankreich) entwickelt. Das Pasteur-Schema verwendet Fragmente von acht Genen (*dinB*, *trpA*, *trpB*, *pabB*, *putP*, *icdA*, *polB* und *uidA*) zur Einordnung der bakteriellen *E. coli*-Stämme in die phylogenetischen Gruppen (Jaureguy et al., 2008). Es wurde auch ein modifiziertes Schema verwendet, das nur die Fragmente von sechs Genen (*trpA*, *trpB*, *pabB*, *putP*, *icdA* und *polB*) beinhaltet und diesen zuspricht, unter geringem Einfluss von horizontalem Gentransfer zu stehen (Escobar-Paramo et al., 2004; Lecomte et al., 1998; Denamur et al., 2000). Mit den Daten der NGS wurde einerseits das herkömmliche MLST-Schema unter Verwendung von acht Genfragmenten mit dem modifizierten Schema unter

Verwendung von sechs Genfragmenten verglichen. Zudem wurden die Sequenzdaten der vollständigen sechs *Housekeeping*-Gene zur MLST genutzt, um den Einfluss der zusätzlichen genetischen Informationen auf die phylogenetische Analyse der Bakterienstämme zu vergleichen. Die Stammbäume wurden mit den Ergebnissen der Triplex-PCR verglichen, die unter "C.2. Ergebnisse der durchgeführten PCRs und Einteilung in PAI-Gruppen" gezeigt wurden.

Für den Stammbaum wurden die Fragmente der sechs Gene *trpA*, *trpB*, *pabB*, *putP*, *icdA* und *polB* verglichen (Abb. 17). Aneinander gereiht besitzen diese Fragmente eine gesamte Sequenzlänge von 3.045 bp. Die Stämme ordnen sich größtenteils in die phylogenetischen Gruppen ein, die zuvor mit der Triplex-PCR bestimmt wurden. Während die Gruppe B2 am eigenständigsten in Bezug auf die anderen phylogenetischen Gruppen ist, spaltet die Gruppe D sich in drei sichtbare Untergruppen auf. Zudem ist der B1-Stamm ECOR30 in dem Stammbaum innerhalb der phylogenetischen Gruppe D zu finden und die B1-Stämme ECOR70 und ECOR72 bilden eine Untergruppe zwischen den Hauptgruppen A und B1.

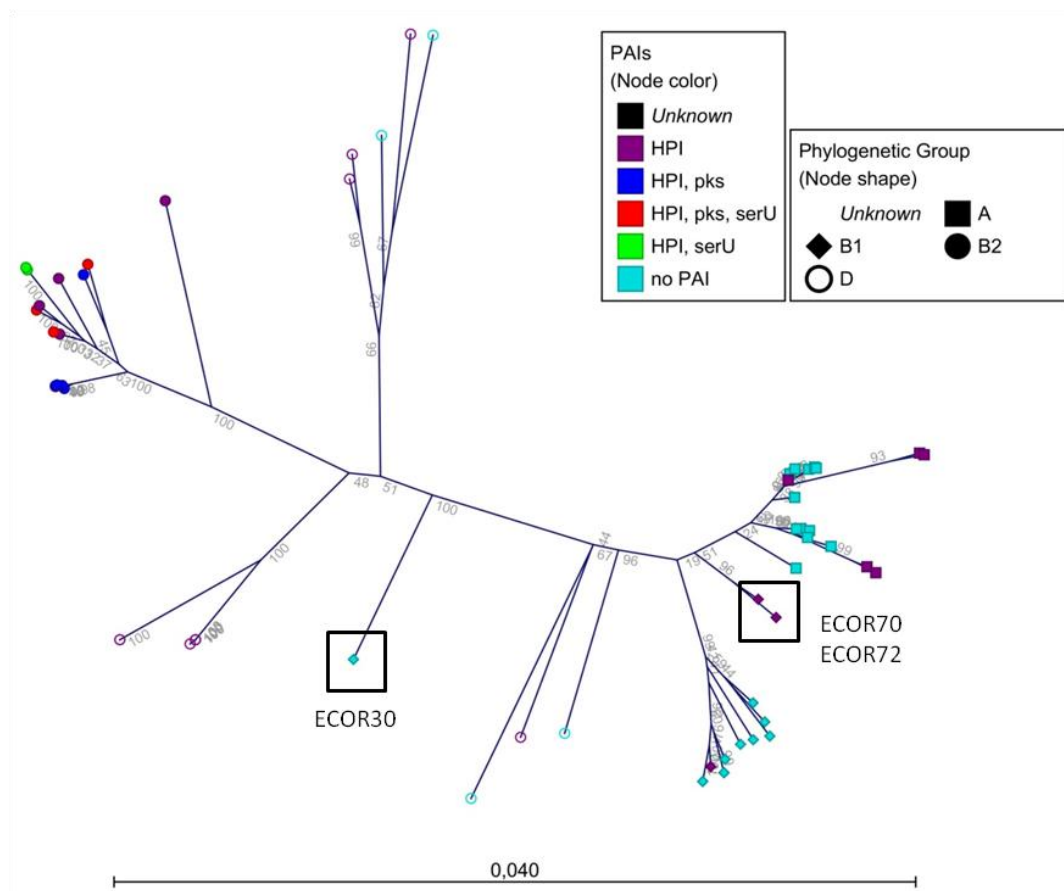


Abbildung 17: Radialer MLST-Stammbaum der *Housekeeping*-Gene. Es wurden die Fragmente der sechs *Housekeeping*-Gene *trpA*, *trpB*, *pabB*, *putP*, *icdA* und *polB* verwendet. Die Legende erläutert die Verteilung der PAIs und der phylogenetischen Gruppen. Der Maßstab zeigt die Anzahl an SNPs pro Base.

Die Fragmente der acht Gene (*dinB*, *trpA*, *trpB*, *pabB*, *putP*, *icdA*, *polB* und *uidA*) haben eine Sequenzlänge von 4.095 bp und bilden den Stammbaum in Abbildung 18. Dieser Baum ist größtenteils identisch mit den Fragmenten der sechs Gene. Die Stämme der phylogenetischen Gruppe B1, ECOR70 und ECOR72, ordnen sich ebenfalls nicht direkt in ihre Gruppe ein. Auch ist der Stamm ECOR30 wieder innerhalb der Gruppe D zu finden. Da in den Genen *dinB* und *uidA* außerhalb der Fragmente große Lücken in den Sequenzen von vielen Isolaten zu finden sind, wurden die acht gesamten Gene nicht als Stammbaum dargestellt und analysiert.

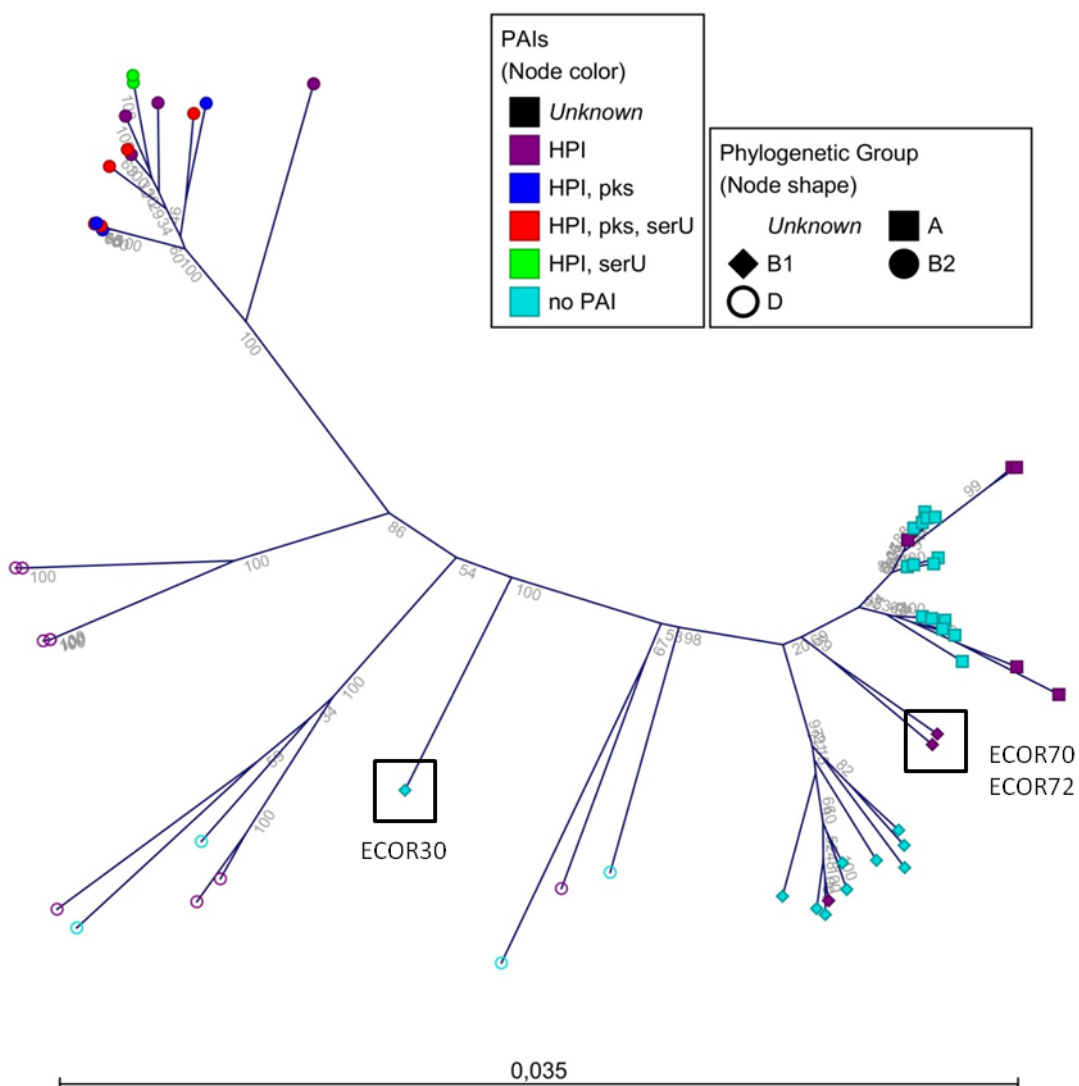


Abbildung 18: Radialer MLST-Stammbaum der *Housekeeping*-Gene. Es wurden die Fragmente der acht Gene *dinB*, *trpA*, *trpB*, *pabB*, *putP*, *icdA*, *polB* und *uidA* verwendet. In der Legende sind die PAI-Gruppen und die phylogenetischen Gruppen erkennbar. Der Maßstab zeigt die Anzahl an SNPs pro Base.

Um zu überprüfen, welche Gene für diese AxB1-Mischgruppe verantwortlich sind, wurde ein *Alignment* mit den betroffenen Stämmen und Vertretern der Gruppe A und B1 erstellt. Es wurde untersucht, in welchen Bereichen außerhalb der für die MLST verwendeten Fragmente die DNA fremder phylogenetischer Gruppen zu finden ist. Das schematische Ergebnis ist in Abbildung 20 zu sehen. Hierfür wurde untersucht, wo sich vermehrt SNPs befinden, die sonst nur innerhalb der anderen phylogenetischen Gruppe vorhanden sind. Das *Alignment* zeigte, dass innerhalb der Fragmente nur DNA der eigenen Gruppe zu finden ist. Die Gene *trpA* und *trpB* sind in allen vier Stämmen nicht von einer Durchmischung mit DNA der anderen phylogenetischen Gruppe betroffen. (i) Bei den beiden Stämmen ECOR16 und ECOR24 der Gruppe A konnten in den Genen *pabB* und *putP* vermehrt DNA der phylogenetischen Gruppe B1 gefunden werden. Bei diesen beiden Isolaten ist die Vermischung mit fremder DNA sehr ähnlich lokalisiert. (ii) Die Stämme ECOR69 und ECOR70 unterscheiden sich größtenteils in dieser Hinsicht. Die Lokalisation von DNA der phylogenetischen Gruppe A in dem Gen *icdA* ist bei beiden Isolaten identisch. In dem Gen *putP* weisen beide Isolate eine Durchmischung mit DNA der phylogenetischen Gruppe A auf, allerdings an verschiedenen Stellen. Im Gegensatz zum Stamm ECOR69 ist in dem Gen *polB* von ECOR70 noch zusätzlich DNA der Gruppe A zu finden. Alle vier untersuchten Isolate weisen also teilweise Hybridgenome auf.

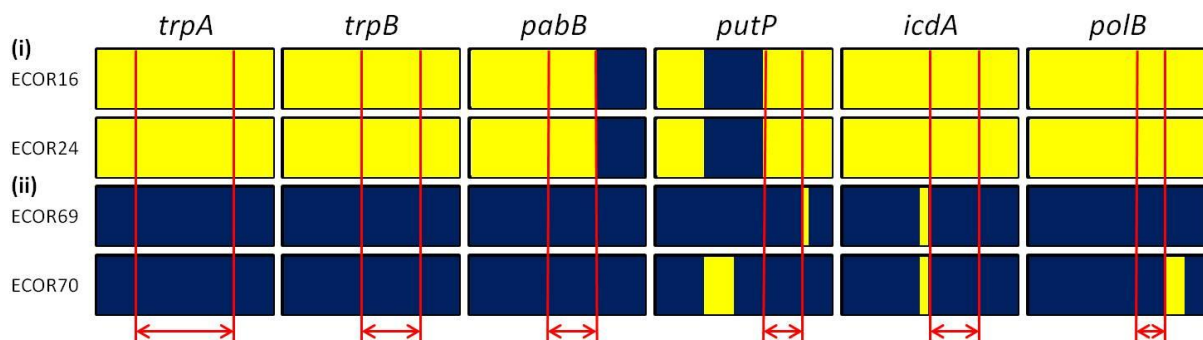


Abbildung 20: Schematische Darstellung der sechs *Housekeeping*-Gene *trpA*, *trpB*, *pabB*, *putP*, *icdA* und *polB* der Stämme ECOR16, ECOR24, ECOR69 und ECOR70. Jedes Rechteck steht für ein Gen. In Rot sind die Fragmente, die für das MLST-Schema verwendet werden, dargestellt. Die Teile von verschiedenen phylogenetischen Abstammungen sind in Gelb (phyl. Gr. A) und in Blau (phyl. Gr. B1) gekennzeichnet.

Zur Untersuchung der evolutionären Geschichte der PAIs wurden erneut die gesamten sechs Gene (*trpA*, *trpB*, *pabB*, *putP*, *icdA* und *polB*) analysiert. Für die Erstellung des Phylogramms in Abbildung 21 wurden nur die Sequenzen von HPI-positiven Stämmen verwendet. Bezüglich der PAI-Gruppen ist zu erkennen, dass die Stämme an sich eine hohe

Verwandtschaft aufzeigen, aber nicht klonal sind. Dies zeigt, dass der PAI-Transfer möglicherweise horizontal und damit *en bloc* stattgefunden hatte, und nicht vertikal. Wäre die Verteilung der spezifischen PAI-Subgruppen durch Zellteilung entstanden, müssten die Stämme der PAI-Gruppen voneinander abstammen.

Für die statistische Auswertung wurden ebenfalls die sechs Gene von allen HPI-positiven Stämmen verwendet. In Tabelle 16 ist zu erkennen, dass die PAI-Gruppe 1 am diversesten ist. Im Vergleich der PAI-Gruppen zeigen die Gruppen 2a und 3 eine höhere Homologie zueinander als zu Gruppe 2b. Innerhalb der phylogenetischen Gruppen ist die Sequenzidentität der Gruppe D am niedrigsten, während die anderen drei Gruppen A, B1 und B2 wesentlich homologer sind. Dies ist bereits in den Stammbäumen ersichtlich.

Der Stamm ECOR65 wurde gesondert behandelt. Dieser zeigt Abweichungen, die nachfolgend erläutert werden.

Tabelle 16: Statistische Analyse der sechs *Housekeeping*-Gene *trpA*, *trpB*, *pabB*, *putP*, *icdA* und *polB*. Ausgewertet wurden alle HPI-positiven Stämme bezüglich der PAI-Gruppe und der phylogenetischen Gruppe. Gezeigt wird die prozentuale Homologie und die durchschnittliche Anzahl an SNPs.

innerhalb der PAI-Gruppen	Homologie	SNPs
Gruppe 2a (n=6)	99,93% (99,88% - 99,98%)	7,5 (2 - 13)
Gruppe 2b (n=1)	99,96%	4
Gruppe 3 (n=15)	99,37% (99,10% - 99,99%)	69,7 (1 - 99)
Gruppe 1 (n=378)	97,68% (96,61% - 100%)	256,7 (0 - 374)
zwischen den PAI-Gruppen	Homologie	SNPs
Vergleich 2b und 2a (n=8)	98,91% (98,88% - 98,95%)	119,6 (116 - 123)
Vergleich 2b und 3 (n=12)	98,96% (98,80% - 99,07%)	114,3 (103 - 132)
Vergleich 2a und 3 (n=24)	99,48% (99,10% - 100%)	57,5 (0 - 99)
zwischen ECOR65 und den PAI-Gruppen	Homologie	SNPs
Vergleich ECOR65 und 2a (n=4)	99,10% (99,08% - 99,13%)	98,8 (96 - 101)
Vergleich ECOR65 und 2b (n=2)	98,89% (98,87% - 98,91%)	122,0 (120 - 124)
Vergleich ECOR65 und 3 (n=6)	99,24% (99,11% - 99,36%)	83,5 (71 - 98)
Vergleich ECOR65 und 1 (n=28)	97,53% (96,90% - 99,36%)	272,1 (71 - 342)
innerhalb der phylogenetischen Gruppen	Homologie	SNPs
Gruppe B2 (n=153)	99,16% (98,11% - 100%)	92,5 (0 - 208)
Gruppe D (n=55)	98,19% (96,85% - 100%)	197,4 (0 - 347)
Gruppe A (n=28)	99,22% (98,51% - 99,98%)	130,9 (2 - 164)
Gruppe B1 (n=6)	99,12% (98,65% - 100%)	96,7 (0 - 149)

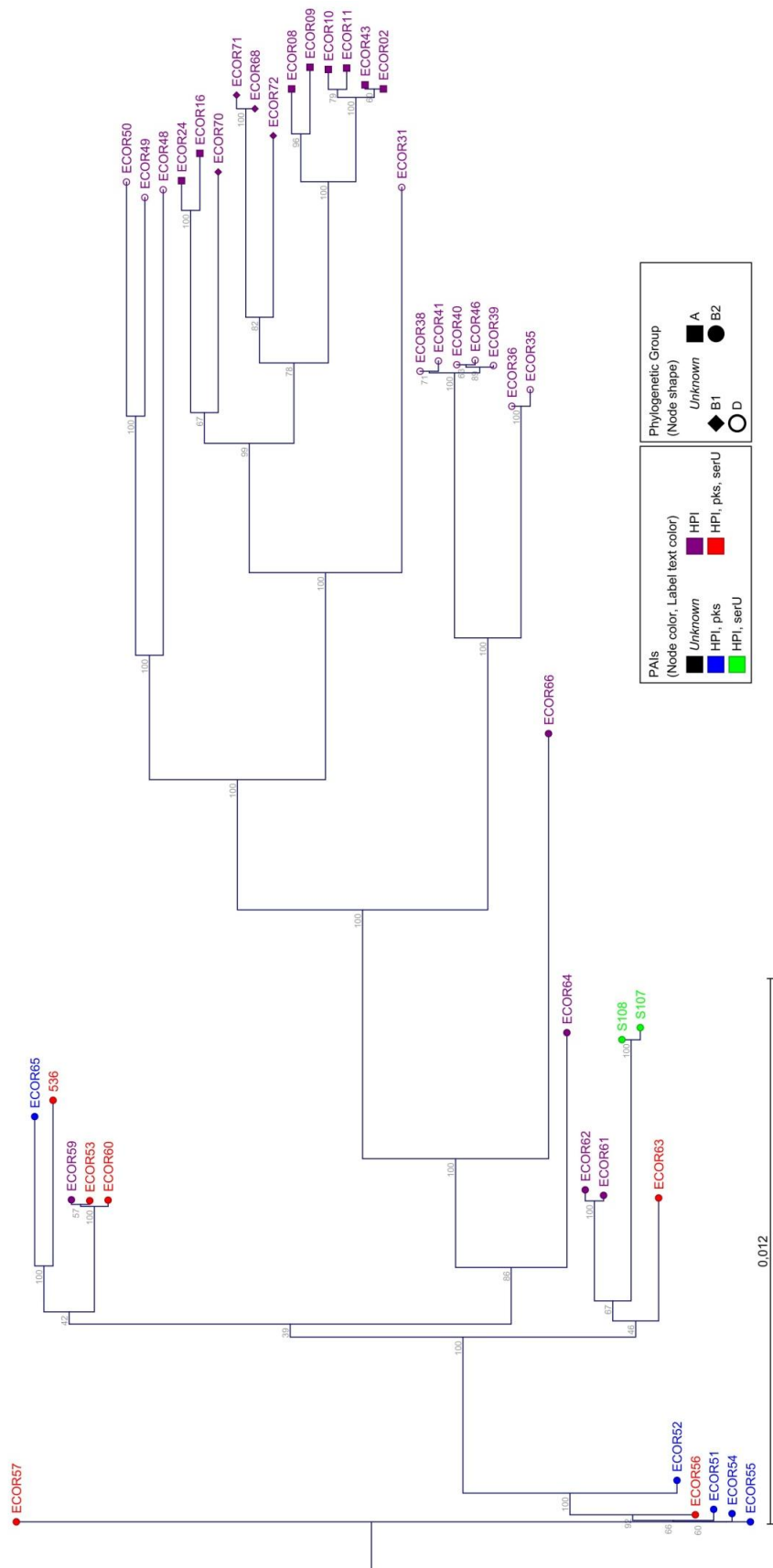


Abbildung 21: MLST-Stammbaum der sechs kompletten *Housekeeping*-Gene aller HPI-positiven Stämme. In der Legende sind die verschiedenen PAI-Gruppen und phylogenetischen Gruppen zu sehen. Der Maßstab beschreibt die Anzahl an SNPs pro Base.

3.2. High-Pathogenicity Island (HPI)

Die HPI codiert für die Synthese und den Transport des Siderophors Yersiniabaktin, das Fe^{3+} -Ionen binden und aus der Umgebung aufnehmen kann (Rakin et al., 1994). Um die evolutionäre Geschichte der HPI zu untersuchen, wurden zwei phylogenetische Stammbäume angefertigt. In dem ersten Stammbaum wird die komplette Insel (ca. 31,5 kb) der untersuchten Bakterienstämme verglichen (Abb. 22). Der Zweite umfasst die sechs Gene *intB*, *ybtQ*, *ybtA*, *irp2*, *irp1* und *fyuA* mit einer gesamten Sequenzlänge von 27.317 bp (Abb. 23). Für den Stammbaum der ganzen Inseln wurde die Sequenz der ICE-Typ-HPI des Stammes ECOR31 auf die Länge der kurzen HPI angepasst, um einen Vergleich mit den HPIs der anderen ECOR-Stämme ziehen zu können.

Bei der Gegenüberstellung der Stammbäume - ganze Insel und MLST - konnten keine großen Unterschiede bezüglich der Verwandtschaft der HPI innerhalb der untersuchten Stämme festgestellt werden. Allerdings entspricht die Sequenzlänge der ausgewählten Gene einem großen Anteil der gesamten Insel.

Die beiden Stammbäume zeigen, dass die Stämme größtenteils entsprechend ihrer phylogenetischen Gruppe angeordnet werden, mit Vermischungen zwischen den Gruppen B2 und D und den Gruppen A und B1. Aufgrund des Vorkommens der HPI in allen phylogenetischen Gruppen war zu vermuten, dass diese die älteste der untersuchten PAIs ist. Die HPI-Sequenzen der Stämme mit zusätzlichen *pks*- und *serU*-Inseln bilden eigene Cluster innerhalb der Gruppe B2. Bei genauerer Betrachtung der Stammbäume wird ersichtlich, dass das Eintreffen der *pks*- und der *serU*-Insel in die Art *E. coli* unabhängig voneinander war und mehrfach stattgefunden hatte. Der Erhalt dieser zusätzlichen Inseln führte zur Bildung der PAI-Gruppen, welche immer die HPI beinhalten. Da die Stämme der drei PAI-Gruppen 2a, 2b und 3 jeweils einen für ihre Gruppe spezifischen HPI-Subtyp besitzen und nicht voneinander abstammen, wurden diese Insel-Kombinationen wahrscheinlich nach der Bildung der PAI-Gruppen *en bloc* via horizontalem Gentransfer (HGT) an andere Bakterien weitergegeben. Die Existenz der spezifischen Subtypen deutet auch darauf hin, dass die Bildung der PAI-Gruppen nicht auf den Verlust von einzelnen Inseln zurückzuführen ist, da die Gruppen sonst denselben oder einen sehr eng verwandten HPI-Subtyp besitzen würden. Eine weitere interessante Beobachtung bei der Betrachtung der phylogenetischen Bäume ist, dass ein Stamm der PAI-Gruppe 2a, ECOR65, nicht wie die anderen Stämme dieser Gruppe angeordnet wird.

Die statistische Auswertung wurde für den Stammbaum der ganzen Insel angefertigt. Hierfür wurde die Homologie und die Anzahl an SNPs berechnet. Die Ergebnisse sind in Tabelle 17 zu sehen. Der Stamm ECOR65 wurde für die Berechnungen nicht in die PAI-Gruppe 2a eingeschlossen, sondern separat mit den Gruppen verglichen. Innerhalb der einzelnen PAI-Gruppen sind die HPIs der Stämme mit Werten zwischen 99,98% und 99,99% nahezu identisch, während die Homologie der Inseln zwischen den PAI-Gruppen (99,54% - 99,63%) immer noch hoch ist. Zudem konnte auch gezeigt werden, dass die HPIs der verschiedenen PAI-Gruppen auch unterschiedliche Sequenzlängen besitzen, was erneut auf spezifische Subtypen hinweist. Die HPI des Stammes ECOR65 hat mit Werten zwischen 99,53% und 99,67% eine ähnliche Homologie zu allen drei PAI-Gruppen. Beim Vergleich aller HPI-Sequenzen, außer der des Stammes ECOR31 mit einer ICE-Typ-HPI, wurde eine prozentuale Identität von 99,53% mit 149,7 SNPs ermittelt.

Tabelle 17: Statistische Auswertung der HPI. Berechnet wurden die Homologie und die Anzahl der SNPs. Es wurden die HPIs innerhalb der einzelnen PAI-Gruppen, zwischen den verschiedenen Gruppen und zwischen dem Stamm ECOR65 und den PAI-Gruppen miteinander verglichen.

innerhalb der PAI-Gruppen	Homologie	SNPs
Gruppe 2a (n=6)	99,99% (99,97% - 100%)	4,5 (1 - 8)
Gruppe 2b (n=1)	99,98%	7
Gruppe 3 (n=15)	99,99% (99,99% - 100%)	2,4 (0 - 4)
zwischen den PAI-Gruppen	Homologie	SNPs
Vergleich 2b und 2a (n=8)	99,63% (99,61% - 99,64%)	116,8 (114 - 122)
Vergleich 2b und 3 (n=12)	99,54% (99,53% - 99,55%)	144,2 (141 - 148)
Vergleich 2a und 3 (n=24)	99,59% (99,57% - 99,60%)	129,9 (126 - 136)
zwischen ECOR65 und den PAI-Gruppen	Homologie	SNPs
Vergleich ECOR65 und 2a (n=4)	99,66% (99,65% - 99,67%)	107,3 (105 - 111)
Vergleich ECOR65 und 2b (n=2)	99,67% (99,66% - 99,67%)	104,5 (103 - 106)
Vergleich ECOR65 und 3 (n=6)	99,53% (99,52% - 99,53%)	149,3 (148 - 151)

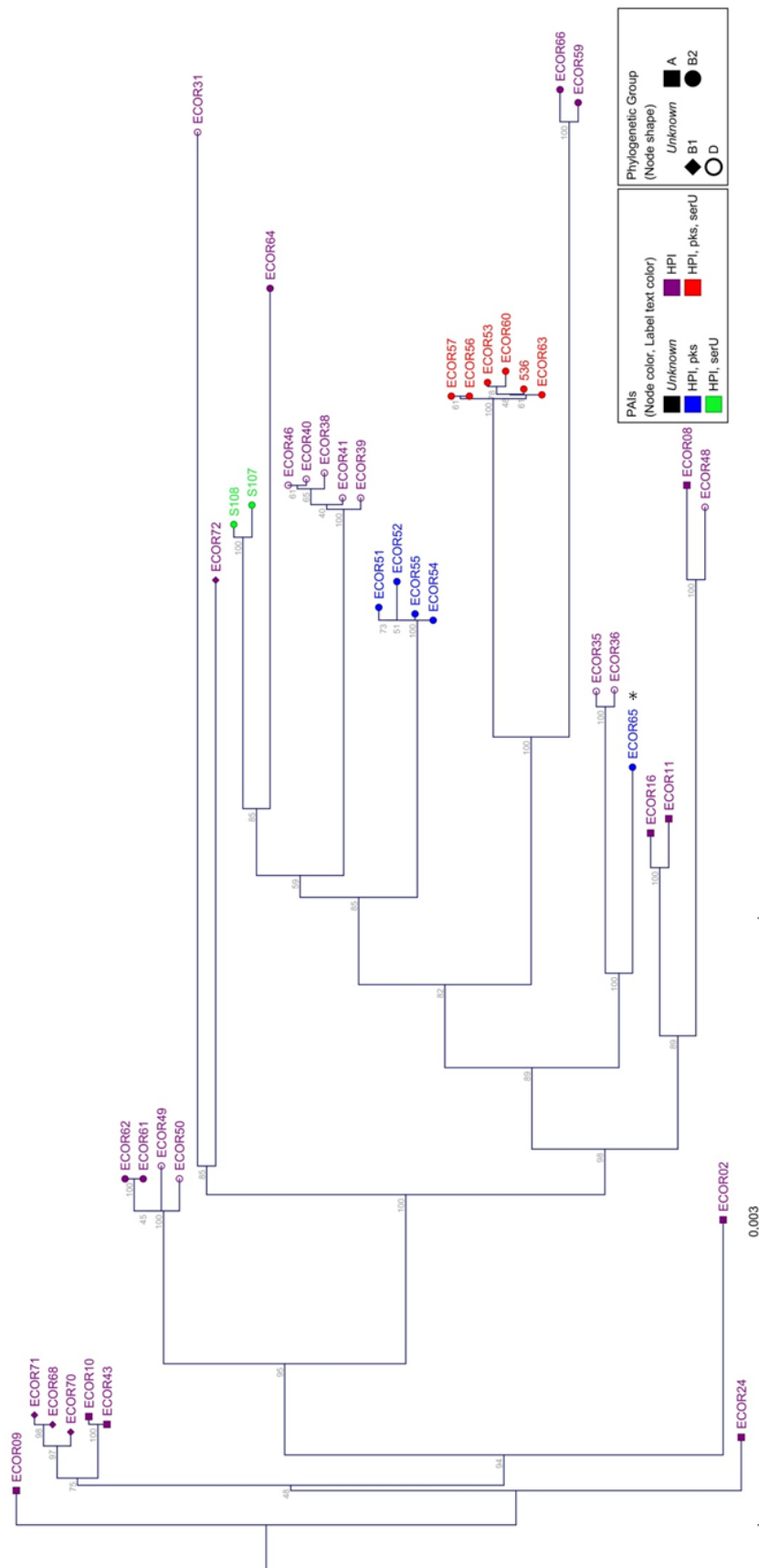


Abbildung 22: Stammbaum der gesamten HPI. In der Legende sind der Farb-Code für die PAI-Gruppen und die Symbole der phylogenetischen Gruppen zu sehen. Der Stamm ECOR65 ist durch einen Stern markiert. Der Maßstab beschreibt die Häufigkeit an SNPs pro Base.

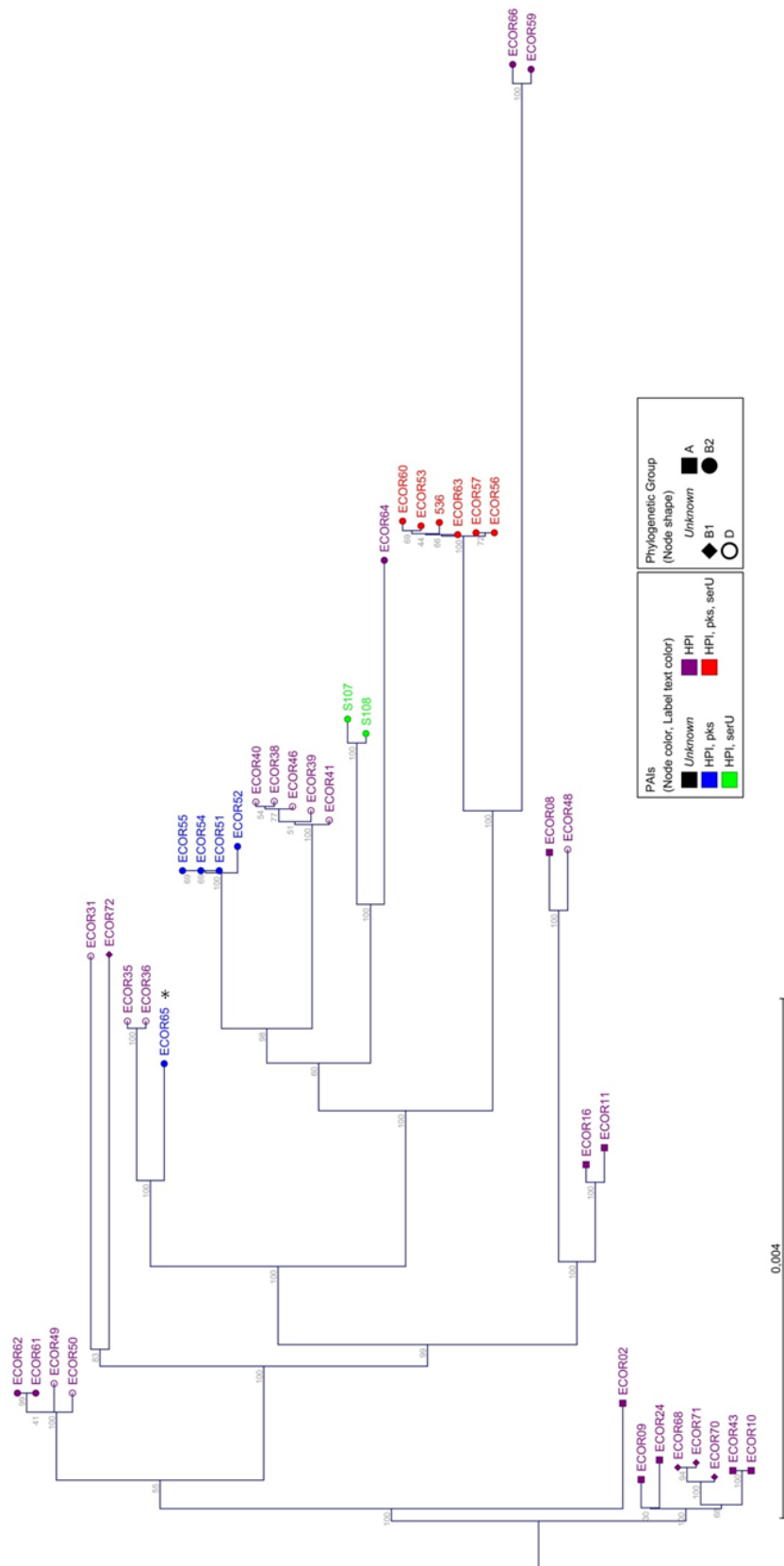


Abbildung 23: MLST-Stammbaum der HPI. Es wurden die sechs Gene *intB*, *ybtQ*, *ybtA*, *irp2*, *irp1* und *fyuA* verwendet. Die Legende beschreibt den Farb-Code der PAI-Gruppen und die Symbole der phylogenetischen Gruppe. Der Stamm ECOR65 ist durch einen Stern markiert. Der Maßstab zeigt die Anzahl an SNPs pro Base.

Um zu überprüfen, wo sich die SNPs innerhalb der HPI des Stammes ECOR65 befinden, wurde ein *Alignment* mit den HPIs der Stämme der PAI-Gruppe 2a angefertigt. Dieses wurde in 3 kb große Abschnitte geteilt, statistisch analysiert und graphisch dargestellt. Die Ergebnisse sind in Abbildung 24 zu sehen. Die SNPs sind größtenteils auf die gesamte HPI verteilt, was für einen unterschiedlichen Subtypen spricht. Die meisten Unterschiede zwischen der ECOR65-HPI und den HPIs der PAI-Gruppe 2a sind in dem Abschnitt 12 - 15 kb zu finden. Bei genauerer Analyse konnten von den durchschnittlichen 23,5 SNPs in einer Sequenzlänge von 121 bp 18 SNPs identifiziert werden. Diese Anhäufung liegt im hinteren Drittel des Gens *irp2*. Außerdem ist die ECOR65-HPI 4 bp kürzer als die der anderen Stämme der PAI-Gruppe 2a. Dafür hat diese PAI dieselbe Sequenzlänge wie die HPIs der PAI-Gruppe 3, deren Sequenzlängen der HPI ebenfalls 4 bp kürzer sind als die der PAI-Gruppe 2a. Allerdings hat die ECOR65-HPI im Vergleich zu den HPIs der PAI-Gruppe 3 nur eine Homologie von 99.53% (Tab. 17).

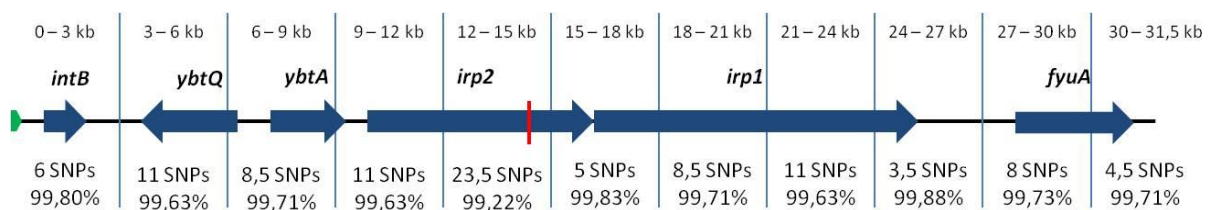


Abbildung 24: Graphische Darstellung der HPI des Stammes ECOR65. Die Insel wurde in 10 Abschnitte à 3 kb und einen Abschnitt à 1,5 kb aufgeteilt und getrennt ausgewertet. Es wurde jeweils die durchschnittliche Anzahl an SNPs und die prozentuale Homologie berechnet. Die *asn* tRNA ist in Grün und die sechs für das MLST verwendeten Gene sind in Blau dargestellt. Die rote Markierung zeigt die Häufung der SNPs innerhalb des Gens *irp2*.

3.3. *pks*-Insel

Die *pks*-Insel codiert für die Synthese des Colibaktins, das Doppelstrangbrüche in der DNA von eukaryotischen Zellen hervorrufen kann (Nougayrede et al., 2006). Für die Phylogenie der *pks*-Insel sollten zwei Stammbäume erstellt werden, einer mit der kompletten Insel (ca. 54,5 kb) und einer mit den sechs Genen *clbA*, *clbE*, *clbI*, *clbL*, *clbO* und *clbQ* (gesamte Sequenzlänge: 8.664 bp). Da innerhalb der *pks*-Insel-positiven Stämme alle sechs Gene zu 100% identisch sind, konnte kein MLST-Stammbaum erstellt werden. Der phylogenetische Baum der gesamten Insel ist in Abbildung 25 zu sehen. In diesem ist zu erkennen, dass die PAI-Gruppen 2a und 3 einen für sich spezifischen *pks*-Insel-Subtyp besitzen. Dies spricht erneut für einen *en bloc*-Transfer. Allerdings besitzt der zur PAI-Gruppe 2a zugehörige

Stamm ECOR65 einen PAI-Subtyp, der sich innerhalb der Gruppe 3 im Baum eingliedert. Der Maßstab des Stammbaumes zeigt ebenfalls, dass die *pks*-Insel allgemein sehr homolog ist. Dies deutet auf eine relativ junge PAI innerhalb der Art *E. coli* hin.

Zwischen den Genen *clbB* und *clbR* existiert eine variable Anzahl an Tandem-Wiederholungen (VNTR; *variable numbers of tandem repeats*) mit unbekannter Funktion, welche in Abbildung 7 dargestellt ist (Putze et al., 2009). Die Anzahl an Tandem-Wiederholungen hat keinen Zusammenhang mit den PAI-Gruppen. Um zu überprüfen, ob die VNTRs eine Auswirkung auf den phylogenetischen Baum haben, wurden diese aus allen *pks*-Insel-Sequenzen *in silico* entfernt. Es konnte keine Veränderung des Stammbaums und somit keine Korrelation zu den PAI-Gruppen festgestellt werden (Daten nicht dargestellt).

Es wurden für die gesamte *pks*-Insel statistische Analysen durchgeführt, deren Ergebnisse in Tabelle 18 zu sehen sind. Innerhalb und zwischen den PAI-Gruppen sind die durchschnittlichen Homologien mit Werten zwischen 99,97% und 99,99% sehr hoch, was das "junge" Alter dieser Insel bestätigt. Die *pks*-Insel des Stammes ECOR65 hat im Vergleich zu denen der PAI-Gruppen 2a und 3 jeweils dieselbe Homologie von 99,89%, ordnet sich im Stammbaum aber innerhalb der PAI-Gruppe 3 an.

Tabelle 18: Statistische Auswertung der *pks*-Insel. Berechnet wurde die Homologie und die Anzahl der SNPs. Es wurden die gesamten *pks*-Inseln innerhalb der PAI-Gruppen 2a und 3, zwischen den beiden Gruppen und zwischen dem Stamm ECOR65 und den PAI-Gruppen mit einander verglichen.

innerhalb der PAI-Gruppen	Homologie	SNPs
Gruppe 2a (n=6)	99,99% (99,98% - 100%)	5,7 (2 - 10)
Gruppe 3 (n=15)	99,97% (99,93% - 100%)	18,3 (2 - 36)
zwischen den PAI-Gruppen	Homologie	SNPs
Vergleich 2a und 3 (n=24)	99,97% (99,93% - 99,99%)	16,0 (4 - 40)
zwischen ECOR65 und den PAI-Gruppen	Homologie	SNPs
Vergleich ECOR65 und 2a (n=4)	99,89% (99,88% - 99,89%)	61,8 (60 - 66)
Vergleich ECOR65 und 3 (n=6)	99,89% (99,87% - 99,90%)	59,3 (55 - 69)

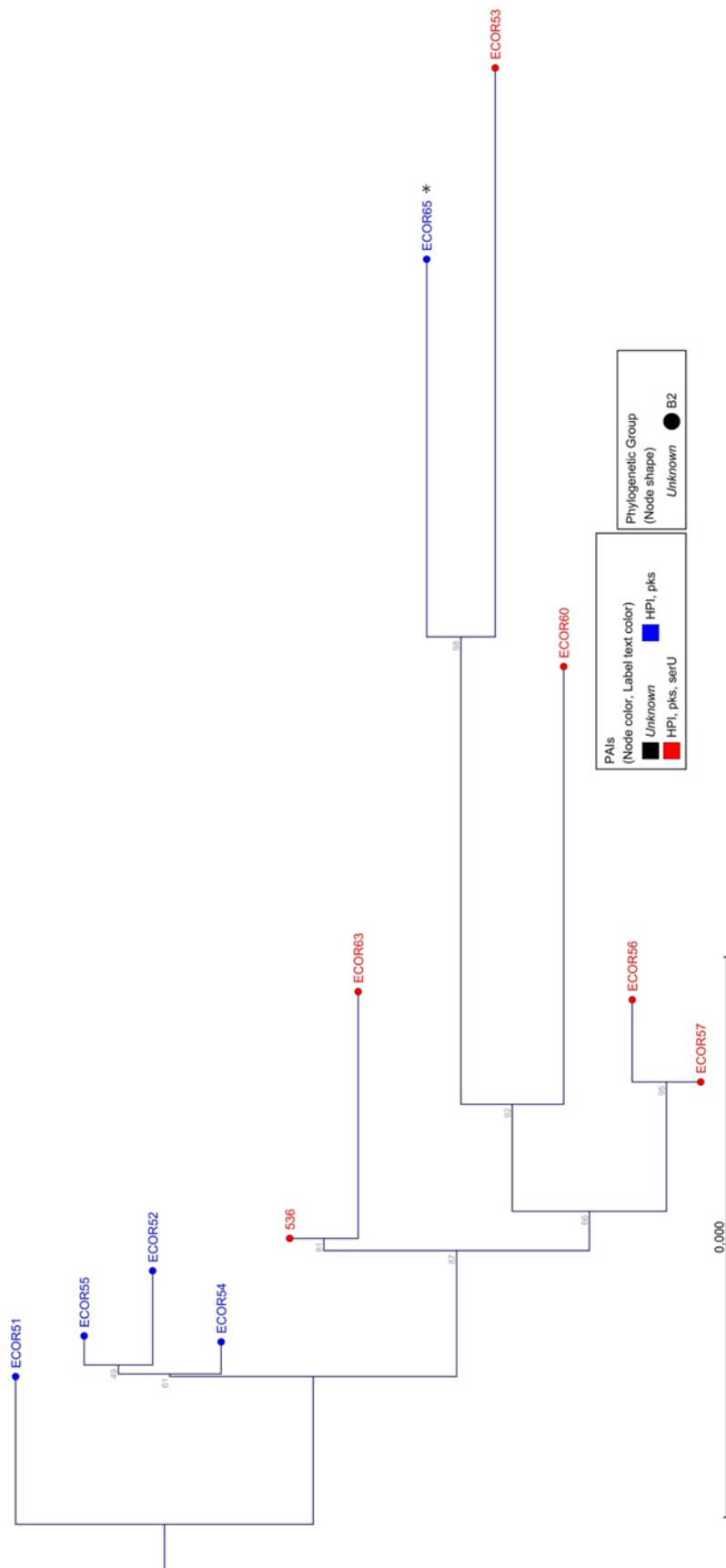


Abbildung 25: Stammbaum der gesamten *pks*-Insel aller positiven Stämme. In der Legende sind die PAI-Gruppen und die phylogenetische Gruppe zu sehen. Der Stamm ECOR65 ist durch einen Stern markiert. Der Maßstab zeigt die Anzahl an SNPs pro Base.

3.4. *serU*-Insel

Die *serU*-Insel codiert für die Synthese des Proteins TcpC in *E. coli*. Aufgrund der Homologie zu den eukaryotischen Motiven der TIR-Domäne von Toll-like Rezeptoren kann TcpC die angeborene Immunantwort des Wirts untergraben (Cirl et al., 2008). Für die Untersuchung der *serU*-Insel wurden zwei phylogenetische Bäume erstellt. Beim Vergleich des Stammbaums der gesamten *serU*-Insel (ca. 27 kb, Abb. 26) mit dem MLST-Stammbaum der sechs Gene *c2395*, *c2398*, *c2400*, *c2403*, *c2408* und *c2414* (gesamte Sequenzlänge: 3.687 bp, Abb. 27) konnte kein großer Unterschied festgestellt werden. Der Stammbaum der gesamten PAI zeigt eine detailliertere Anordnung der Sequenzen, die vor allem innerhalb der PAI-Gruppe 3 zu sehen ist. Dies kommt durch die Unterschiede außerhalb der sechs untersuchten Gene der Insel zustande. In beiden Stammbäumen sind zwei *serU*-Insel-Subtypen zu sehen, einer der PAI-Gruppe 3 und einer der Gruppe 2b. Das Vorkommen der zwei verschiedenen Subtypen ergänzt die Ergebnisse der HPI und der *pks*-Insel und ist ein weiteres Indiz für einen *en bloc*-Transfer.

Die statistische Auswertung der gesamten *serU*-Insel kann eindeutig belegen, dass es sich um unterschiedliche Subtypen handelt (Tab. 19). Die Sequenz-Homologie der Inseln innerhalb der Gruppe 2b liegt bei 99,93%, die der Gruppe 3 bei einem ähnlichen Wert von 99,94%. Zwischen den PAI-Gruppen wurde eine viel geringere Homologie mit einem Wert von 99,27% errechnet. Die Unterschiede zwischen den Gruppen sind auch deutlich in den phylogenetischen Stammbäumen zu erkennen.

Tabelle 19: Statistische Auswertung der *serU*-Insel. Dargestellt wurde die Homologie und die Anzahl der SNPs. Es wurden die gesamten *serU*-Inseln innerhalb der PAI-Gruppen 2b und 3 und zwischen den beiden Gruppen verglichen.

innerhalb der PAI-Gruppen	Homologie	SNPs
Gruppe 2b (n=1)	99,93%	18
Gruppe 3 (n=15)	99,94% (99,86% - 99,99%)	16,6 (2 - 39)
zwischen den PAI-Gruppen	Homologie	SNPs
Vergleich 2b und 3 (n=12)	99,27% (99,20% - 99,30%)	197,3 (189 - 215)

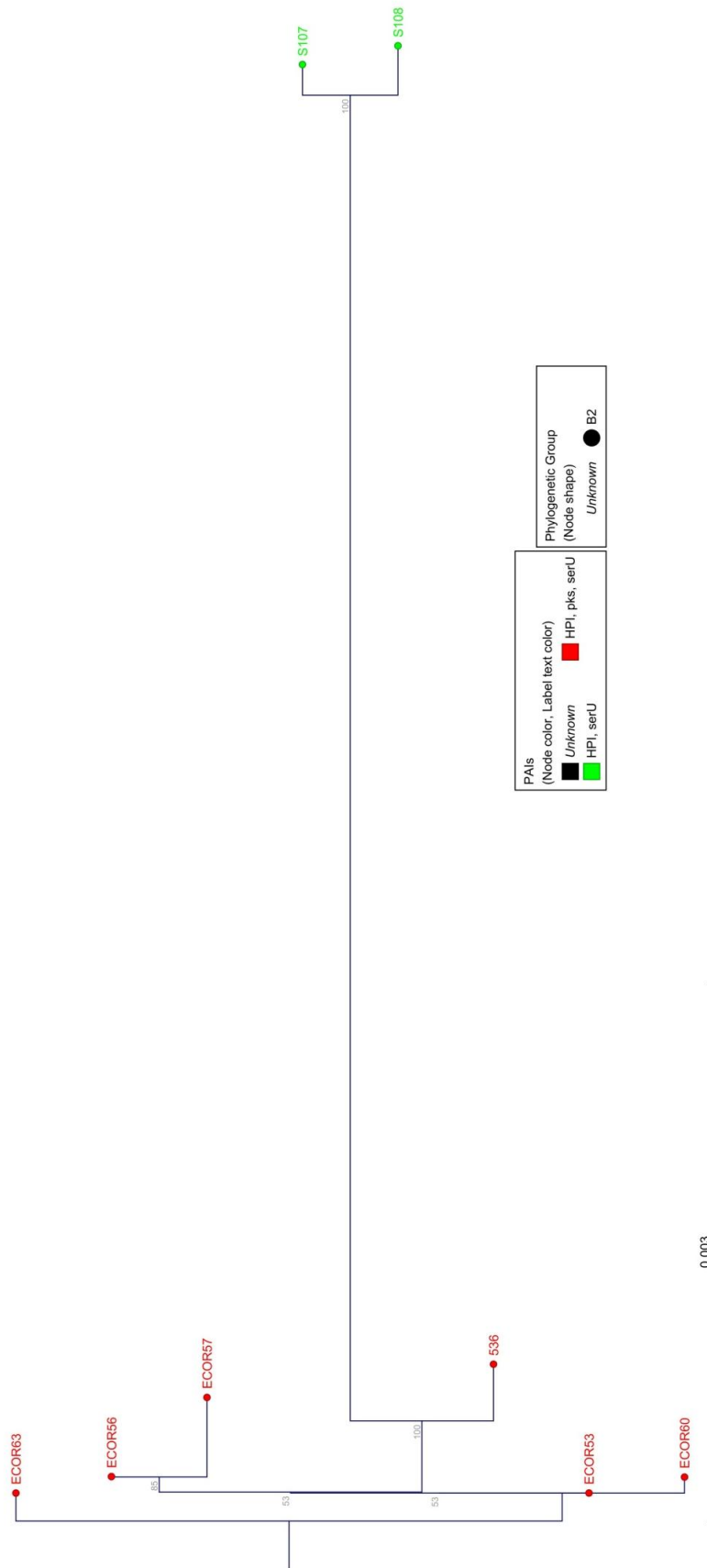


Abbildung 26: Stammbaum der gesamten *serU*-Insel aller positiven Stämme. Die Legende zeigt die unterschiedlichen PAI-Gruppen und die phylogenetische Gruppe. Der Maßstab beschreibt die Anzahl an SNPs pro Base.

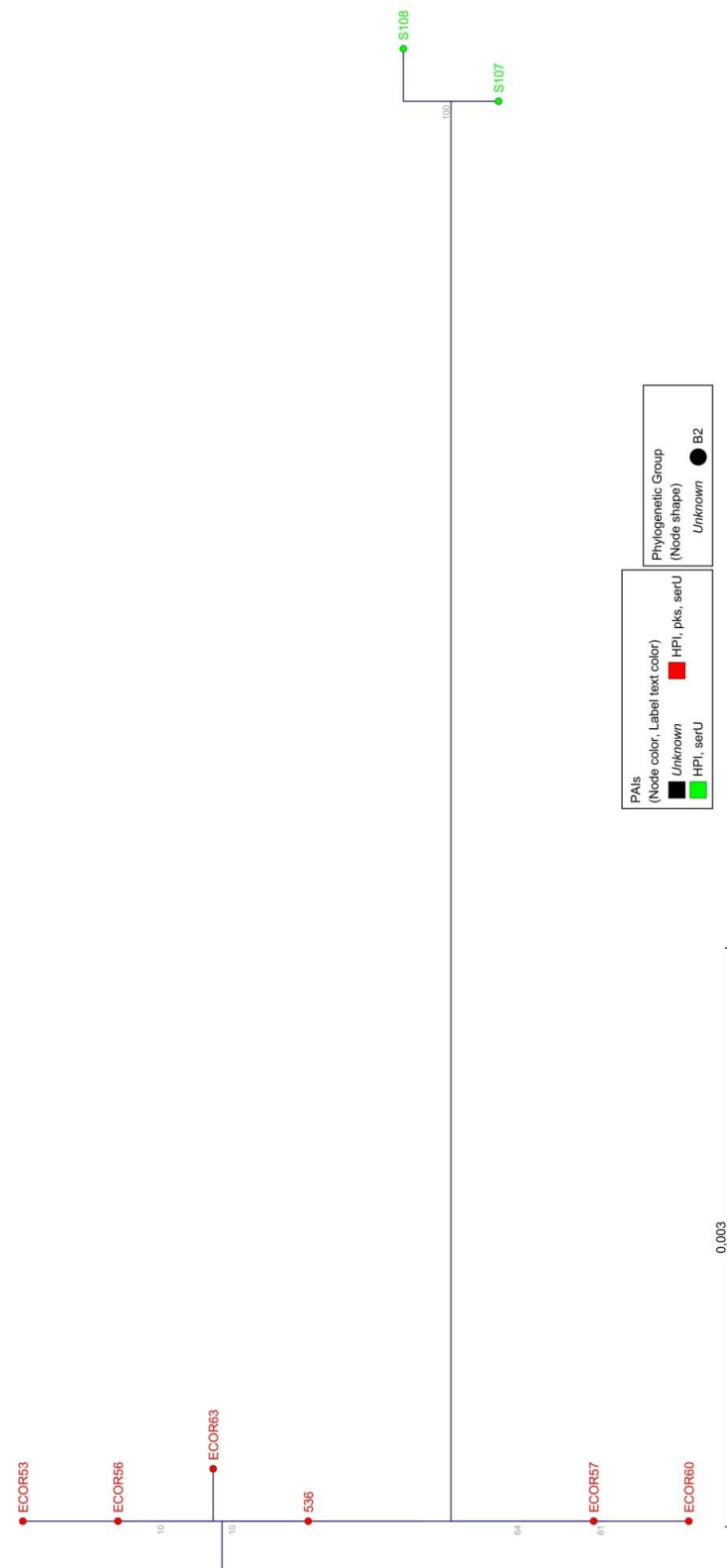


Abbildung 27: MLST-Stammbaum der *serU*-Insel. Es wurden die sechs Gene *c2395*, *c2398*, *c2400*, *c2403*, *c2408* und *c2414* verwendet. In der Legende sind die PAI-Gruppen und die phylogenetische Gruppe zu sehen. Der Maßstab zeigt die Anzahl an SNPs pro Base.

3.5. Inter-PAI-Regionen

Die Theorie des *en bloc*-Transfers besagt, dass die gesamten Inseln innerhalb eines Transfers an einen Rezipienten weitergegeben wurden. Da für die einzelnen Inseln für jede PAI-Gruppe immer spezifische Subtypen gefunden wurden, sollten nun die DNA-Abschnitte zwischen den Inseln untersucht werden. Diese zwei Regionen wurden als inter-PAI-Regionen bezeichnet, eine Region zwischen der *serU*-Insel und der HPI, und eine zwischen der HPI und der *pks*-Insel.

3.5.1. Inter-PAI-Region *serU* - HPI

Die inter-PAI-Region zwischen der *serU*-Insel und der HPI hat eine Sequenzlänge von ca. 1 kb. Der zugehörige Stammbaum ist in Abbildung 28 zu sehen. Es ist zu erkennen, dass die verschiedenen PAI-Gruppen auch sehr homologe Bereiche zwischen den beiden Inseln besitzen. Der Stamm ECOR65 der PAI-Gruppe 2a ist nicht in seiner Gruppe zu finden. Die anderen ECOR-Stämme gruppieren sich weniger nach ihrer phylogenetischen Gruppe ein, sondern vielmehr nach dem Vorhandensein der HPI. Zwei HPI-positive Stämme, ECOR31 und ECOR72, ordnen sich zu den HPI-negativen Stämmen ein. Die statistische Auswertung bestätigt die durch den Stammbaum erlangten Erkenntnisse (Tab. 20). Innerhalb der PAI-Gruppen 2a und 2b ist diese inter-PAI-Region zu 100% homolog, innerhalb der Gruppe 3 zu 99,97%. Zwischen den PAI-Gruppen ist die Homologie mit Werten zwischen 99,60% bis 99,78% viel geringer.

Tabelle 20: Statistische Auswertung der inter-PAI-Region zwischen der *serU*-Insel und der HPI. Es wurden die prozentuale Homologie und die Anzahl an SNPs innerhalb der PAI-Gruppen, zwischen den PAI-Gruppen und zwischen dem Stamm ECOR65 und den PAI-Gruppen berechnet.

innerhalb der PAI-Gruppen	Homologie	SNPs
Gruppe 2a (n=6)	100% (100% - 100%)	0 (0 - 0)
Gruppe 2b (n=1)	100%	0
Gruppe 3 (n=15)	99,97% (99,90% - 100%)	0,3 (0 - 1)
zwischen den PAI-Gruppen	Homologie	SNPs
Vergleich 2b und 2a (n=8)	99,60% (99,60% - 99,60%)	4,0 (4 - 4)
Vergleich 2b und 3 (n=12)	99,78% (99,70% - 99,80%)	2,2 (2 - 3)
Vergleich 2a und 3 (n=24)	99,78% (99,70% - 99,80%)	2,2 (2 - 3)
zwischen ECOR65 und den PAI-Gruppen	Homologie	SNPs
Vergleich ECOR65 und 2a (n=4)	99,70% (99,70% - 99,70%)	3,0 (3 - 3)
Vergleich ECOR65 und 2b (n=2)	99,70% (99,70% - 99,70%)	3,0 (3 - 3)
Vergleich ECOR65 und 3 (n=6)	99,88% (99,80% - 99,90%)	1,2 (1 - 2)

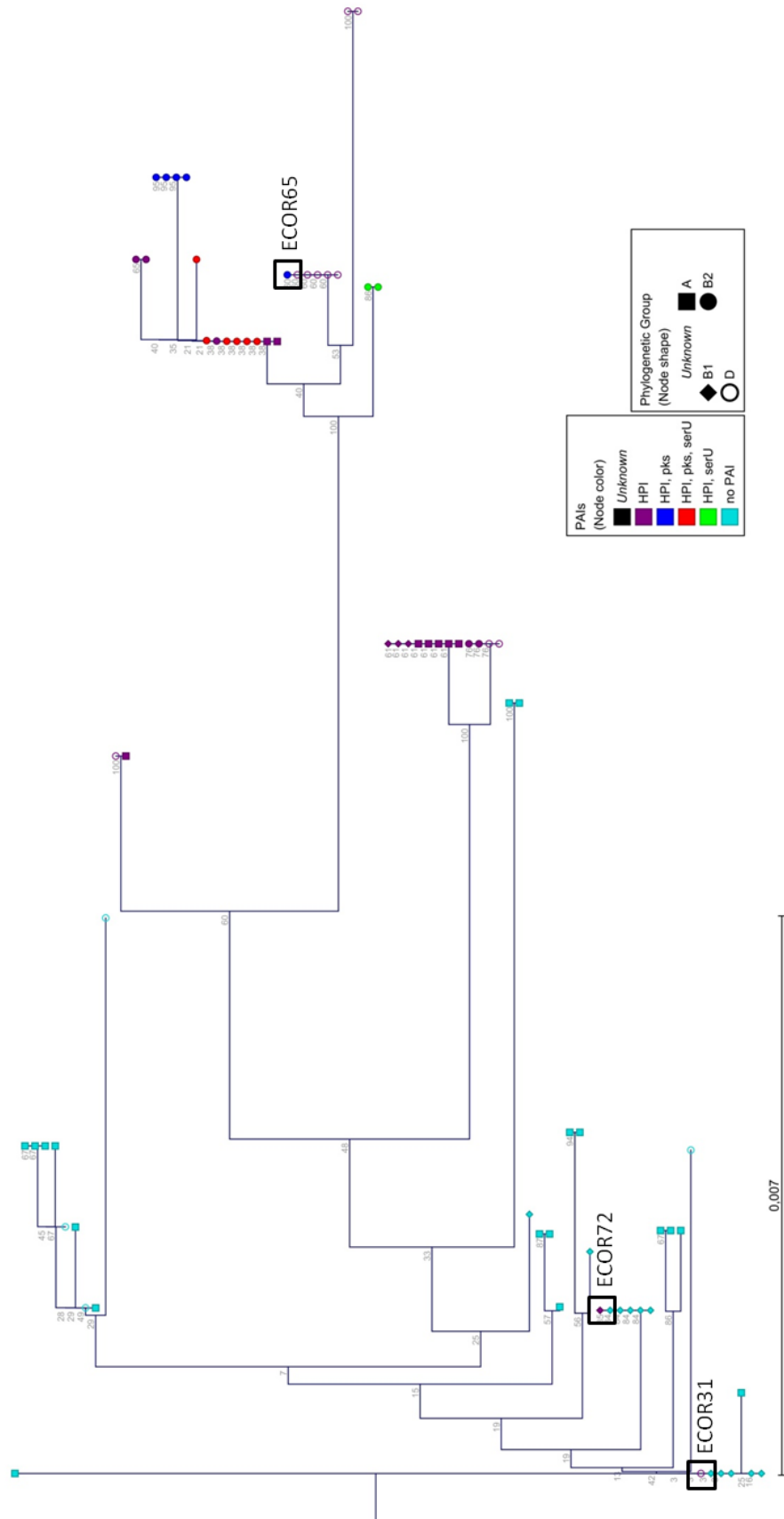


Abbildung 28: Stammbaum der inter-PAI-Region zwischen der *serU*-Insel und der HPI. Die Stämme ECOR31, ECOR65 und ECOR72 sind gekennzeichnet. Die Legende beschreibt die PAI-Gruppe und die phylogenetischen Gruppen. Der Maßstab zeigt die Anzahl an SNPs pro Base.

3.5.2. Inter-PAI-Region HPI - pks

Die inter-PAI-Region zwischen der HPI und der *pks*-Insel hat eine Länge von ca. 12,5 kb. Der Stammbaum der für diesen Abschnitt angefertigt wurde ist in Abbildung 29 zu sehen. Zu erkennen ist, dass diese genomische Region innerhalb der einzelnen PAI-Gruppen ebenfalls sehr homolog ist. Auch hier ordnen sich die übrigen ECOR-Stämme eher nach dem Vorhandensein der HPI als ihrer phylogenetischen Gruppe ein, außer den Stämmen ECOR31 und ECOR72.

Es wurde eine statistische Auswertung angefertigt, die in Tabelle 21 zu sehen ist. Zu erkennen ist eine sehr hohe Homologie von 99,98% bis 99,99% innerhalb der einzelnen PAI-Gruppen. Beim Vergleich zwischen den einzelnen Gruppen sind geringere Übereinstimmungen erkennbar, vor allem zwischen den Gruppen 2b und 2a und den Gruppen 2b und 3 mit Werten von 99,28% und 99,31%. Homologer ist die inter-PAI-Region im Vergleich zwischen den PAI-Gruppen 2a und 3 mit einem Wert von 99,87%. Die Region des Stamms ECOR65 hat die höchste Homologie zur Gruppe 3 (99,97%). Die PAI-Gruppe 2a, in die dieser Stamm eigentlich einzuordnen ist, hatte im Vergleich eine Homologie von 99,86%. Der geringste Wert ist im Vergleich zu Gruppe 2b (99,31%) zu finden.

Da sich innerhalb der PAI-Gruppen beide inter-PAI-Regionen phylogenetisch wie die entsprechenden Inseln verhalten, konnte die Theorie des *en bloc*-Transfers *in silico* bestätigt werden.

Tabelle 21: Statistische Auswertung der inter-PAI-Region zwischen der HPI und der *pks*-Insel. Es wurden die prozentuale Homologie und die Anzahl an SNPs innerhalb der PAI-Gruppen, zwischen den PAI-Gruppen und zwischen dem Stamm ECOR65 und den PAI-Gruppen berechnet.

innerhalb der PAI-Gruppen	Homologie	SNPs
Gruppe 2a (n=6)	99,99% (99,98% - 99,99%)	1,3 (1 - 2)
Gruppe 2b (n=1)	99,98%	3
Gruppe 3 (n=15)	99,98% (99,96% - 100%)	2,3 (0 - 5)
zwischen den PAI-Gruppen	Homologie	SNPs
Vergleich 2b und 2a (n=8)	99,28% (99,26% - 99,30%)	90,5 (88 - 93)
Vergleich 2b und 3 (n=12)	99,31% (99,28% - 99,33%)	86,8 (85 - 91)
Vergleich 2a und 3 (n=24)	99,87% (99,84% - 99,88%)	17,0 (15 - 20)
zwischen ECOR65 und den PAI-Gruppen	Homologie	SNPs
Vergleich ECOR65 und 2a (n=4)	99,86% (99,85% - 99,87%)	18,0 (17 - 19)
Vergleich ECOR65 und 2b (n=2)	99,31% (99,29% - 99,32%)	87,5 (86 - 89)
Vergleich ECOR65 und 3 (n=6)	99,97% (99,95% - 99,98%)	3,3 (2 - 6)



Abbildung 29: Stammbaum der inter-PAI-Region zwischen der HPI und der *pks*-Ins. Die Stämme ECOR31 und ECOR72 sind hervorgehoben. In der Legende sind die PAI-Gruppen und die phylogenetischen Gruppen zu sehen. Der Maßstab beschreibt die Anzahl an SNPs pro Base.

3.6. PAI III

Die PAI III von *E. coli* codiert für die Synthese des Siderophors Salmochelin (Russo et al., 1999). Zur Untersuchung der PAI III wurden mehrere Stammbäume erstellt. Um einen möglichen *en bloc*-Transfer zusammen mit den bereits untersuchten PAIs zu ermitteln, wurden ebenfalls die Genombereiche *upstream* und *downstream* der PAI III untersucht. Hierfür wurde von jedem Stamm eine Sequenz von 10 kb verwendet. Da in den Sequenzierungen teilweise große Lücken in der Region um die PAI III und auf der Insel selber vorhanden waren, konnten nicht alle Sequenzen der Bakterienstämme genutzt werden. Innerhalb der PAI III war aufgrund fehlender Sequenzen nur der Vergleich einer 10,5 kb großen Region möglich, die die Gene *iroB*, *iroC*, *iroD*, *iroE* und *iroN* enthielten. Der *iroBCDEN* Gen-Cluster der PAI III codiert für die Synthese des Rezeptors IroN. Dieser Gen-Cluster wurde auch auf dem *E. coli* Plasmid p300 gefunden (Sorsa et al., 2003).

3.6.1. Teilstück der Insel

Der Stammbaum in Abbildung 31 vergleicht die *iro*-Gene aller PAI III-positiven Stämme. Es wurden auch statistische Berechnungen durchgeführt, die in Tabelle 22 zu sehen sind. Die Sequenzen der Stämme ECOR51, ECOR52, ECOR53, ECOR54, ECOR65 und ECOR60 sind mit 100% Homologie und durchschnittlichen 0,3 SNPs nahezu identisch. Die Stämme ECOR63, ECOR66 und 536 ordnen sich einzeln an, teilen sich aber eine Homologie von 99,36% mit 68,0 SNPs. Es konnte kein Bezug zu den PAI-Gruppen gefunden werden, obwohl die Gruppe 2a komplett identisch bezüglich ihrer Sequenzen ist. Da sich der Cluster der Stämme ECOR62 und ECOR70 weit entfernt von den anderen einreicht, wurde nach anderen bekannten *iroBCDEN* Gen-Clustern gesucht. Die bereits veröffentlichte Sequenz des *E. coli*-Plasmids p300 (NCBI-Zugangsnummer NG_035855), das ebenfalls die *iro*-Gene enthält, wurde in den Stammbaum eingefügt. Es konnte gezeigt werden, dass der auf 10,5 kb getrimmte Bereich des Plasmids sehr homolog zu dem Cluster der zwei ECOR-Sequenzen war. Deshalb wurde vermutet, dass sich der *iroBCDEN* Gen-Cluster dieser Stämme auf einem Plasmid befindet. Es konnte auch *in silico* nachgewiesen werden, dass in dem Genom der Stämme ECOR62 und ECOR70 keine PAI III integriert ist. Als Beispiel wird der Abschnitt des ECOR70-Genoms in Abbildung 30 gezeigt.

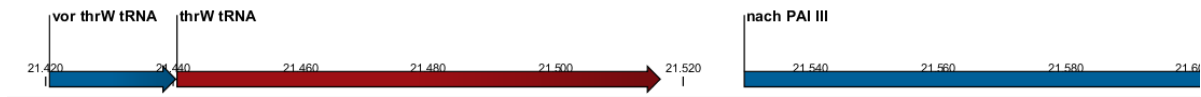


Abbildung 30: Ausschnitt aus dem Genom des Stammes ECOR70. Gezeigt wird die *thrW* tRNA (rot) und deren Umgebung. In blau dargestellt sind die Bereiche vor und nach der PAI III.

Zwischen der *thrW* tRNA und dem Bereich nach der PAI III ist im Stamm ECOR70 eine Sequenz von 13 bp zu finden, während sich beim Stamm ECOR62 lediglich zwei Basen befinden. Ein BLAST-Vergleich war aufgrund der kurzen Sequenz nicht möglich.

Die genomische PAI III hat eine Homologie von 99,69% mit durchschnittlich 33,1 SNPs und die episomale Insel eine Ähnlichkeit von 99,86% mit 14,7 SNPs. Beim Vergleich der beiden PAI III-Typen ist lediglich eine Homologie von 98,14% mit 196,6 SNPs zu finden. Der Stamm ECOR65 wurde nicht in die statistische Analyse der PAI-Gruppen eingeschlossen, allerdings bei der Berechnung der genomischen PAI verwendet. Die Sequenz des Plasmids p300 wurde für die Statistik der episomalen Insel mit einbezogen.

Tabelle 22: Statistische Auswertung eines 10,5 kb großen Abschnitts innerhalb der PAI III. Es wurde die Homologie und die Anzahl an SNPs berechnet. Für die Berechnungen der PAI-Gruppen wurde der Stamm ECOR65 ausgeschlossen, allerdings wurde er in der Berechnung der genomischen PAI berücksichtigt.

innerhalb der PAI-Gruppen	Homologie	SNPs
Gruppe 1 (n=3)	98,80% (98,28% - 99,83%)	126,7 (18 - 182)
Gruppe 2a (n=3)	100% (100% - 100%)	0 (0 - 0)
Gruppe 3 (n=6)	99,61% (99,41% - 99,99%)	40,8 (1 - 62)
zwischen den PAI-Gruppen	Homologie	SNPs
Vergleich 1 und 2a (n=9)	98,49% (98,05% - 99,31%)	159,7 (88 - 93)
Vergleich 1 und 3 (n=12)	98,54% (98,04% - 99,45%)	154,5 (58 - 207)
Vergleich 2a und 3 (n=12)	99,78% (99,57% - 100%)	22,8 (0 - 45)
genomische und episomale PAI III	Homologie	SNPs
genomische PAI III (n=36)	99,69% (99,21% - 100%)	33,1 (0 - 84)
episomale PAI III (n=3)	99,86% (99,83% - 99,91%)	14,7 (10 - 18)
Vergleich genomisch und episomal (n=27)	98,14% (98,04% - 98,30%)	196,6 (180 - 207)
phylogenetische Gruppen	Homologie	SNPs
innerhalb Gruppe B2 (n=45)	99,37% (98,04% - 100%)	66,4 (0 - 207)
Vergleich zwischen B1 und B2 (n=10)	98,33% (98,10% - 99,83%)	176,9 (18 - 201)
weitere Vergleiche	Homologie	SNPs
ECOR51, ECOR52, ECOR53, ECOR54, ECOR65 und ECOR60 (n=15)	100% (99,99% - 100%)	0,3 (0 - 1)
ECOR63, ECOR66 und 536 (n=3)	99,36% (99,21% - 99,45%)	68,0 (58 - 84)

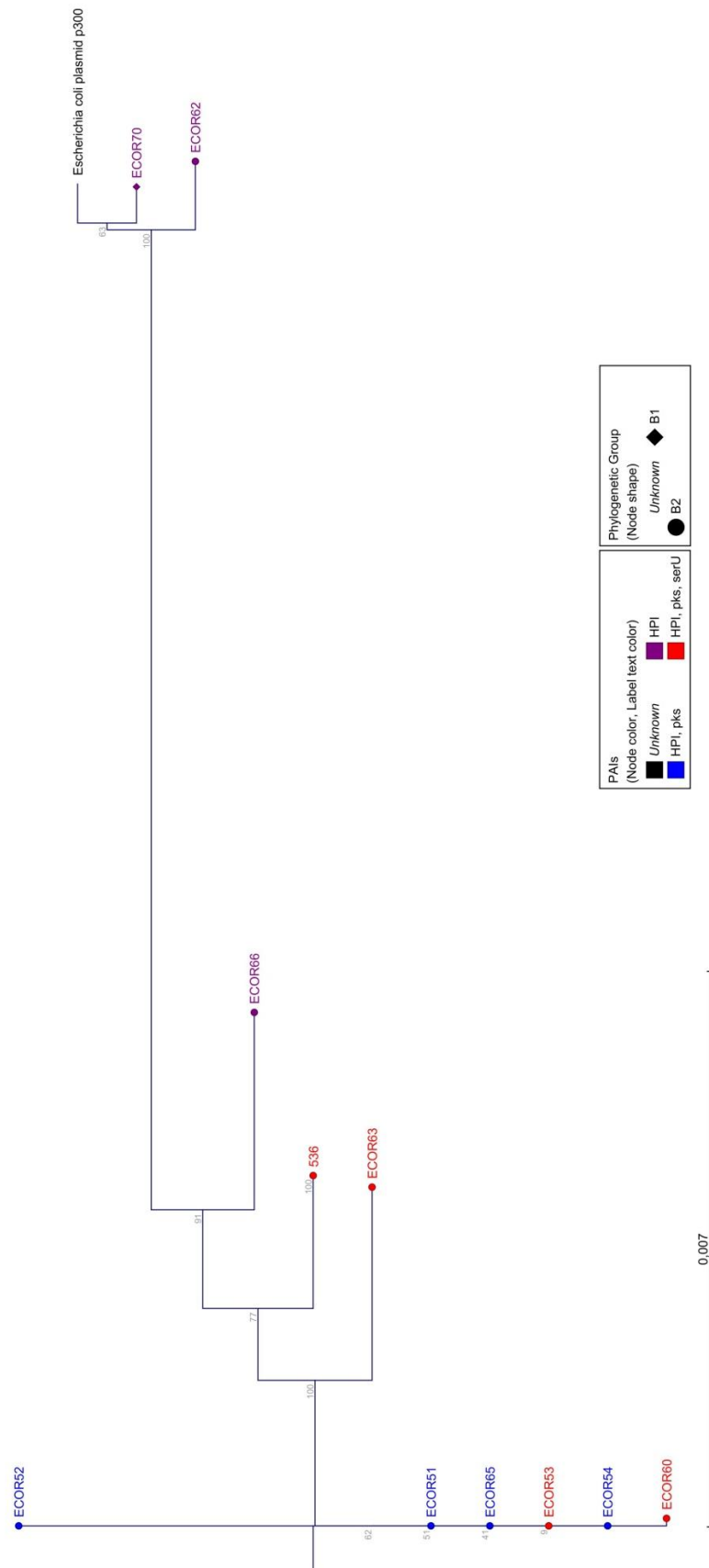


Abbildung 31: Stammbaum eines 10,5 kb großen Bereiches innerhalb der PAI III, der den *iro*-Cluster enthält. Die Sequenz des Plasmid p300 wurde auf die entsprechende Größe getrimmt. Die PAI-Gruppen und die phylogenetischen Gruppen sind in der Legende zu sehen. Der Maßstab gibt die Anzahl an SNPs pro Base an.

3.6.2. Upstream-Bereich

In Abbildung 32 ist der Stammbaum der Sequenzen 10 kb *upstream* der PAI III zu sehen, in Tabelle 23 die dazugehörige statistische Auswertung. Es ist zu erkennen, dass die Sequenzen der PAI-Gruppe 2a, außer der des Stammes ECOR65, identisch sind. Dies war nicht der Fall für die beiden anderen PAI-Gruppen. Während die Gruppe 3 mit 99,47% Homologie und durchschnittlichen 52,7 SNPs noch recht verwandt ist, ist in der Gruppe 1 mit 88,71% und durchschnittlichen 1143,0 SNPs keine große Ähnlichkeit erkennbar. Dies liegt insbesondere an der Sequenz des Stammes ECOR70 bzw. an seiner phylogenetischen Gruppe. Während innerhalb der Gruppe B2 noch eine Homologie von 98,31% mit durchschnittlich 103,2 SNPs zu finden ist, konnte verglichen mit der Gruppe B1 nur noch eine Ähnlichkeit von 84,23% mit 1588,4 SNPs errechnet werden.

Bei genauerer Betrachtung des *Alignments*, aus dem dieser Stammbaum hervorgeht, werden die in dem Baum erkennbaren Unterschiede sichtbar. Hierfür wurden die Abweichungen zwischen den einzelnen Sequenzen der Stämme analysiert. Im Allgemeinen ist die Sequenz des Stammes ECOR70 relativ homolog zu den anderen untersuchten Sequenzen. Allerdings sind in diesem Stamm zwei Deletionen und eine Insertion zu finden. In den anfänglichen 8 kb, ausgehend von der *thrW* tRNA, sind noch keine Abweichungen zu finden. Die erste Deletion von 169 bp ist nach diesen erwähnten 8 kb in dem laut BLAST-Analyse hypothetischen *peptide chain release factor-like protein*, präsent. Noch 500 bp weiter ist ebenfalls in diesem Gen die zweite Deletion von 1027 bp zu sehen. Nach insgesamt 9,6 kb ist eine Insertion von 128 bp zu finden, die laut BLAST-Analyse zu einer Acetyl-Transferase gehört. Somit können 1324 Sequenz-Unterschiede den Deletionen und der Insertion zugeschrieben werden und sind damit nicht nur reine Homologie-Abweichungen.

Bei dem statistischem Vergleich der genomischen und der episomalen PAI III ist zu erkennen, dass der *upstream*-Bereich der genomischen eine viel höhere Homologie als der der episomalen PAI besitzt. Dafür ist vor allem wieder der Stamm ECOR70 bzw. seine phylogenetische Gruppe an den Homologie- und den SNP-Werten beteiligt.

Tabelle 23: Statistische Auswertung des Bereichs 10 kb *upstream* der PAI III. Es wurden die Homologien und die Anzahl an SNPs berechnet. Die Analysen wurden für die Sequenzen der Stämme innerhalb und zwischen den PAI-Gruppen durchgeführt. Die Berechnungen wurden auch mit den Sequenzen der genomischen und der episomalen PAI III sowie mit den unterschiedlichen phylogenetischen Gruppen angefertigt.

innerhalb der PAI-Gruppen	Homologie	SNPs
Gruppe 1 (n=3)	88,71% (83,56% - 98,87%)	1143,0 (113 - 1665)
Gruppe 2a (n=3)	100% (100% - 100%)	0 (0 - 0)
Gruppe 3 (n=6)	99,47% (99,18% - 100%)	52,7 (0 - 81)
zwischen den PAI-Gruppen	Homologie	SNPs
Vergleich 1 und 2a (n=9)	93,67% (84,39% - 98,41%)	636,0 (159 - 1570)
Vergleich 1 und 3 (n=12)	93,72% (84,33% - 98,73%)	631,0 (127 - 1576)
Vergleich 2a und 3 (n=12)	99,28% (99,23% - 99,30%)	71,8 (70 - 76)
genomische und episomale PAI III	Homologie	SNPs
genomische PAI III (n=36)	99,06% (98,21% - 100%)	93,1 (0 - 179)
episomale PAI III (n=1)	83,56%	1665
Vergleich genomisch und episomal (n=18)	91,43% (83,70% - 98,87%)	861,8 (113 - 1651)
phylogenetische Gruppen	Homologie	SNPs
innerhalb Gruppe B2 (n=45)	98,31% (98,21% - 100%)	103,2 (0 - 179)
Vergleich zwischen B1 und B2 (n=10)	84,23% (98,10% - 99,83%)	1588,4 (1560 - 1665)

Um die Frage zu klären, ob die Diversität des Stammes ECOR70 in dem *upstream*-Bereich der PAI III von der phylogenetischen Gruppe abhängig ist, wurden die B1-Stämme der ECOR-Sammlung mit eingeschlossen (Abb. 33). Aufgrund fehlender Sequenzen konnten nicht alle B1-Stämme in den phylogenetischen Baum mit einbezogen werden. Es ist zu erkennen, dass die B1-Stämme verschiedene Cluster bilden, sich aber nicht dem Stamm ECOR70 zuordnen. Der B2-Referenzstamm 536 grenzt sich deutlich von den anderen Stämmen ab. Die Sequenz des Stammes ECOR58 ist in dem phylogenetischen Stammbaum zwischen den Sequenzen der Stämme ECOR70 und 536 zu finden.

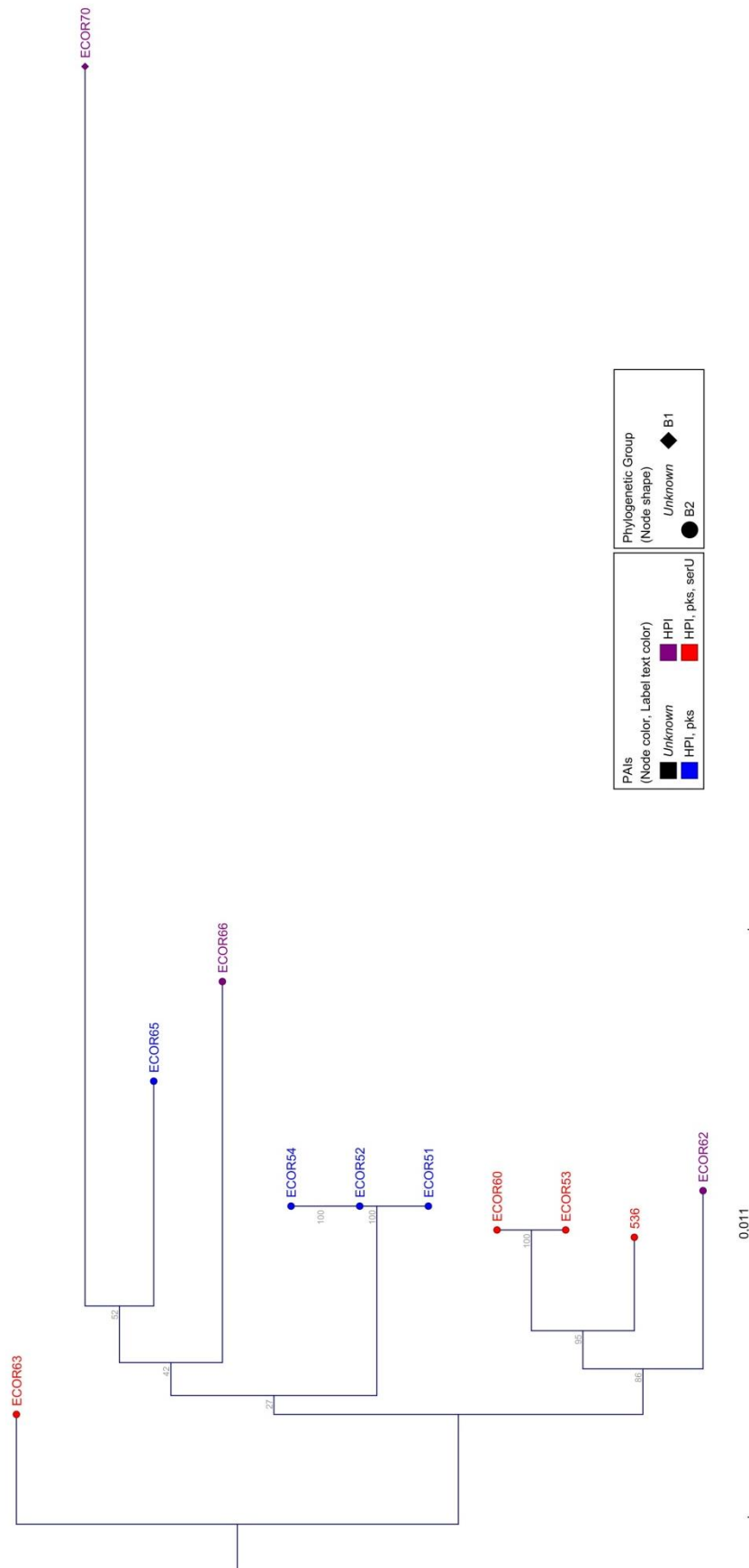


Abbildung 32: Stammbaum des 10 kb langen *upstream*-Bereichs der PAI III. In der Legende werden die PAI-Gruppen und die phylogenetischen Gruppen erläutert. Der Maßstab repräsentiert die Anzahl an SNPs pro Base.

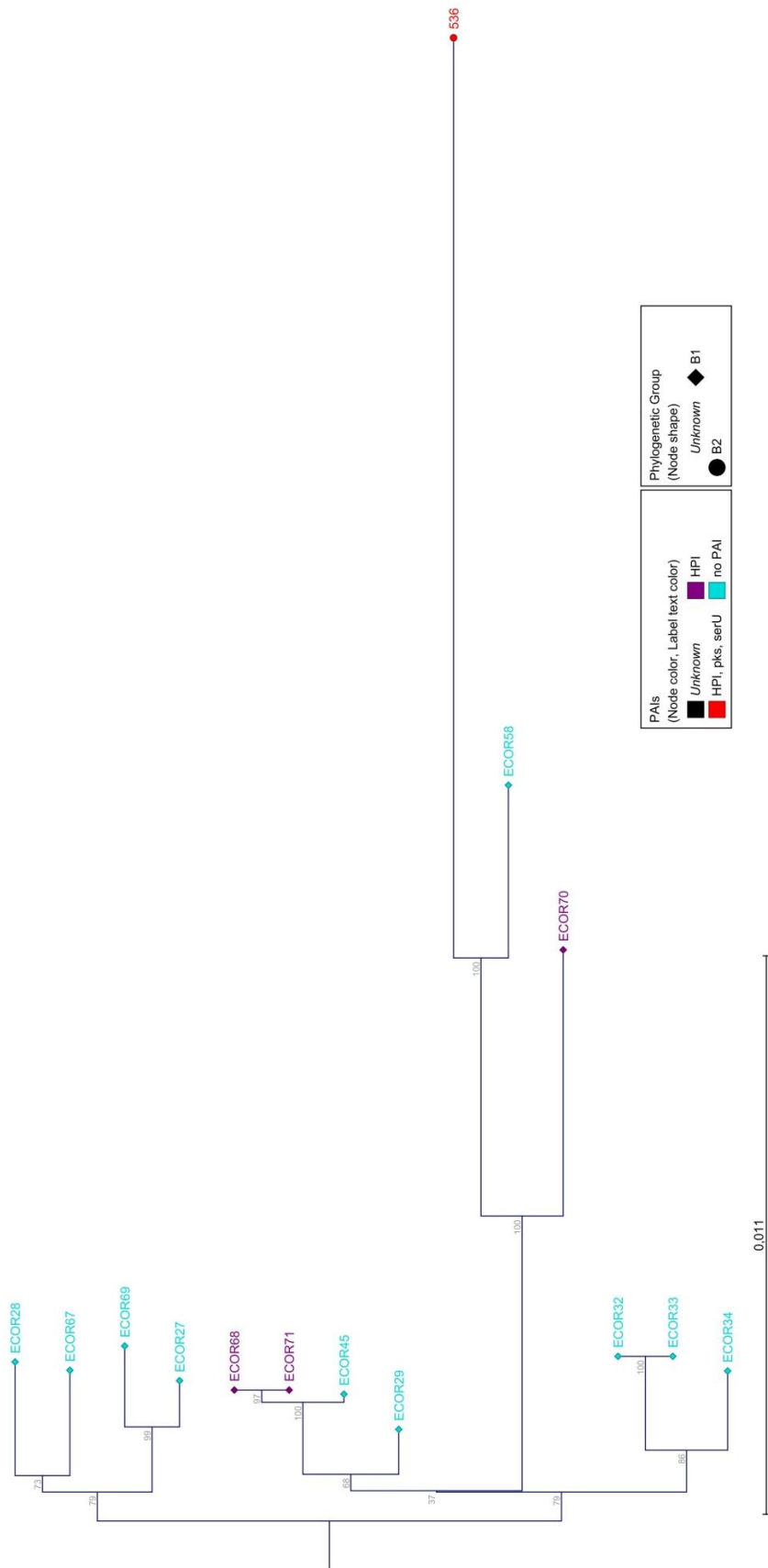


Abbildung 33: Stammbaum der phylogenetischen Gruppe B1 und des Stammes 536. Es wurde der 10 kb lange *upstream*-Bereich der PAI III dargestellt. Die PAI-Gruppen und die phylogenetischen Gruppen sind in der Legende zu sehen. Der Maßstab beschreibt die Anzahl an SNPs pro Base.

3.6.3. Downstream-Bereich

Der in Abbildung 34 gezeigte Stammbaum repräsentiert die 10 kb langen Sequenzen *downstream* der PAI III. Der Stamm ECOR65 ist nicht vertreten, da dessen *downstream*-Sequenz nicht vorhanden ist. Auf den ersten Blick ist erkennbar, dass sich der Stamm ECOR70 massiv von den anderen Stämmen unterscheidet. Die statistische Auswertung, die in Tabelle 24 abgebildet ist, bestätigt dies. Während innerhalb der phylogenetischen Gruppe B2 eine Homologie von 99,58% und durchschnittlich 41,7 SNPs errechnet wurden, ist im Vergleich zur Gruppe B1 nur noch eine sehr geringe Homologie von 42,36% mit 6129,9 SNPs zu finden. Die Unterschiede werden bei Betrachtung des *Alignments*, welches für die Erstellung des Stammbaums generiert wurde, sofort sichtbar. Obwohl die ersten 500 bp noch recht homolog sind, kann man in diesem Abschnitt bereits eine Deletion von 11 bp und eine Insertion von 12 bp finden. Das Ergebnis einer BLAST-Analyse der ersten 180 bp nach diesem Abschnitt ist eine Phagen-Integrase. Anschließend wird die Sequenz immer diverser. Teilweise ist noch ein Vergleich mit den anderen Sequenzen möglich, meist unterscheidet sich der *downstream*-Bereich des Stammes ECOR70 komplett von diesen. Dieser Bereich ist ebenfalls von Insertionen und Deletionen geprägt.

Tabelle 24: Statistische Auswertung der 10 kb langen Sequenzen des *downstream*-Bereichs der PAI III. Der Stamm ECOR65 wurde aufgrund fehlender Sequenzen nicht berücksichtigt. Es wurden die Homologien und die Anzahl an SNPs pro Base berechnet. Hierfür wurden Vergleiche innerhalb und zwischen den PAI-Gruppen gezogen, sowie für die beiden PAI III-Typen und die phylogenetischen Gruppen.

innerhalb der PAI-Gruppen	Homologie	SNPs
Gruppe 1 (n=3)	61,43% (42,38% - 99,52%)	4101,0 (48 - 6128)
Gruppe 2a (n=3)	99,91% (99,87% - 99,98%)	9,3 (2 - 13)
Gruppe 3 (n=6)	99,60% (99,49% - 99,95%)	40,0 (5 - 51)
zwischen den PAI-Gruppen	Homologie	SNPs
Vergleich 1 und 2a (n=9)	80,52% (42,35% - 99,77%)	2070,4 (23 - 6131)
Vergleich 1 und 3 (n=12)	80,50% (42,29% - 99,75%)	2072,3 (25 - 6137)
Vergleich 2a und 3 (n=12)	99,50% (99,32% - 99,72%)	49,8 (28 - 68)
genomische und episomale PAI III	Homologie	SNPs
genomische PAI III (n=28)	99,57% (99,32% - 99,98%)	42,6 (2 - 68)
episomale PAI III (n=1)	42,38%	6128
Vergleich genomisch und episomal (n=16)	70,99% (42,29% - 99,77%)	3084,3 (23 - 6137)
phylogenetische Gruppen	Homologie	SNPs
innerhalb Gruppe B2 (n=36)	99,58% (99,32% - 99,98%)	41,7 (2 - 68)
Vergleich zwischen B1 und B2 (n=9)	42,36% (42,29% - 42,41%)	6129,9 (6124 - 6137)

Zur besseren Einsicht der anderen Stämme wurde ein erneuter Stammbaum ohne ECOR70 erstellt (Abb. 35). Hier ist zu erkennen, dass die PAI-Gruppe 2a eine Art Cluster bildet, der mit 99,91% noch relativ homolog ist. Auch die Gruppe 3 ist mit einer Homologie von 99,60% untereinander näher verwandt, allerdings zeigt der Stammbaum eine Durchmischung mit den beiden Stämmen ECOR62 und ECOR66 aus PAI-Gruppe 1.

Auch für den *downstream*-Bereich der PAI III sollte geklärt werden, ob die Diversität des Stammes ECOR70 von der phylogenetischen Gruppe abhängig ist. Es wurden die B1-Stämme der ECOR-Sammlung mit eingeschlossen, von denen die Sequenz verfügbar ist. Der Stammbaum in Abbildung 36 zeigt, dass die meisten Sequenzen der B1-Stämme nahezu identisch sind, allerdings nicht die von ECOR70. Dieser zeigt eine nähere Verwandtschaft mit ECOR26 und eine entferntere mit ECOR69. Der B2-Stamm 536 ordnet sich zwischen diesen beiden B1-Gruppen ein.

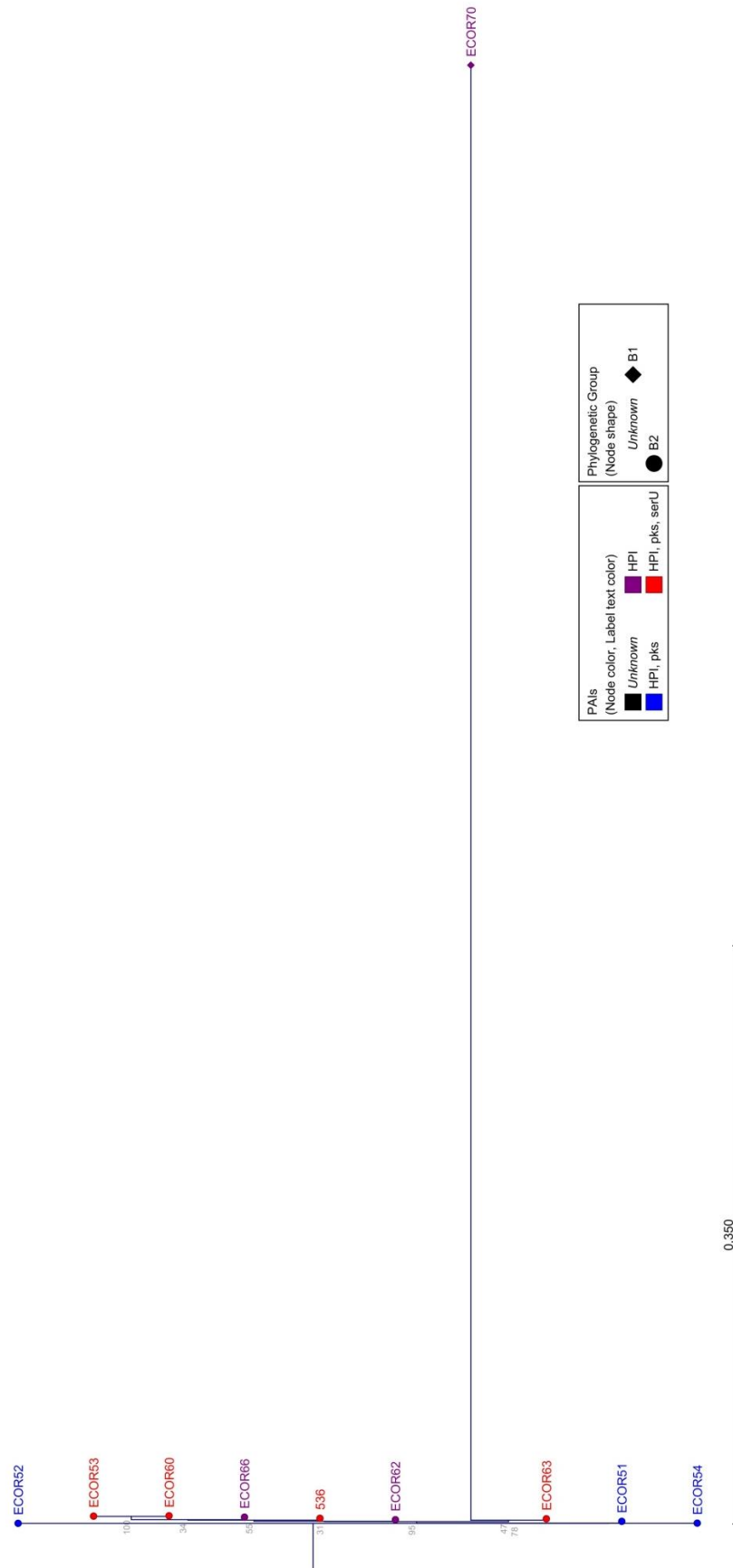


Abbildung 34: Stammbaum der Sequenzen 10 kb *downstream* der PAI III. Die PAI-Gruppen und die phylogenetischen Gruppen sind in der Legende dargestellt. Der Maßstab zeigt die Anzahl der SNPs pro Base.

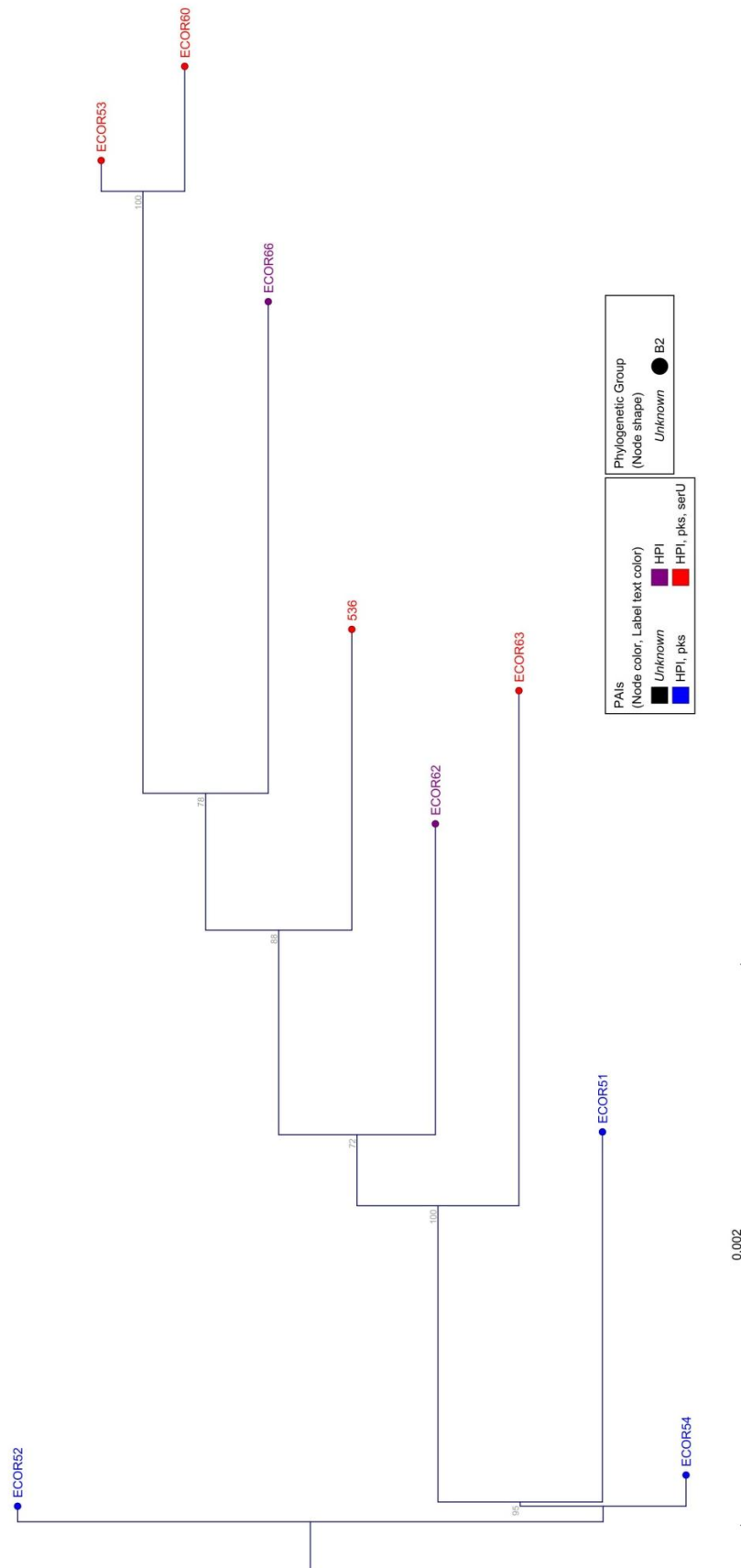


Abbildung 35: Stammbaum der Sequenzen 10 kb *downstream* der PAI III ohne den Stamm ECOR70. In der Legende werden die PAI-Gruppen und die phylogenetischen Gruppen dargestellt. Der Maßstab repräsentiert die Anzahl der SNPs pro Base.

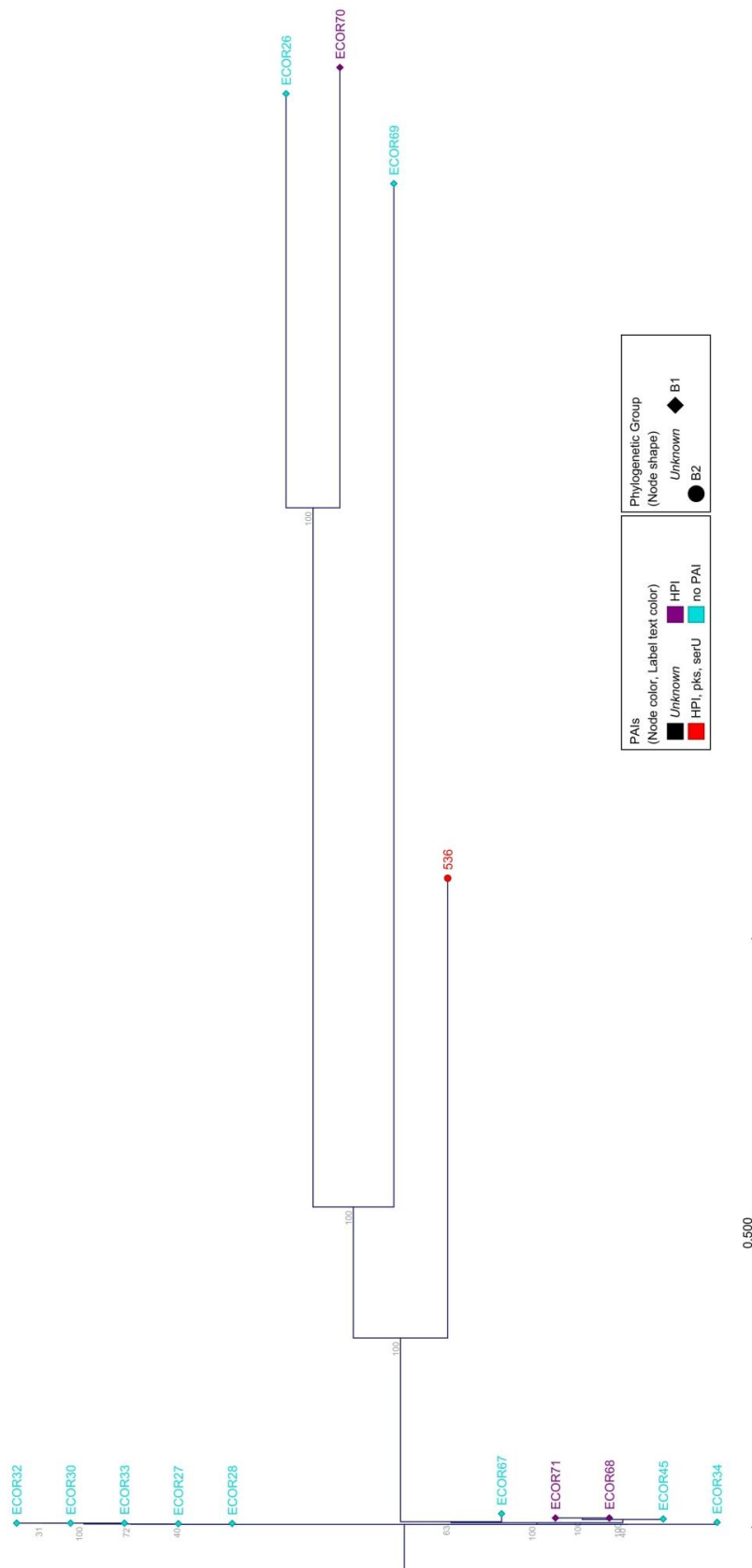


Abbildung 36: Stammbaum des 10 kb langen *downstream*-Bereichs der PAI III. Es wurden die Stämme der phylogenetischen Gruppe B1 und der Stamm 536 verwendet. Die PAI-Gruppen und die phylogenetischen Gruppe sind in der Legende zu sehen. Der Maßstab stellt die Anzahl an SNPs pro Base dar.

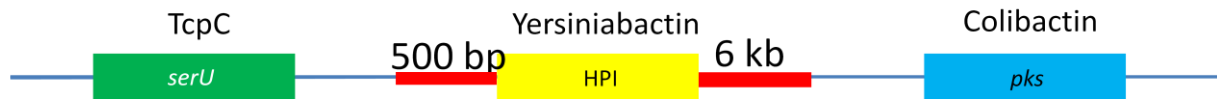
3.7. Untersuchung des Stammes ECOR65

Da der Stamm der PAI-Gruppe 2a, ECOR65, sich phylogenetisch nicht zu den dazugehörigen Gruppenmitgliedern einordnet, wurde dieser genauer untersucht. Oft ist eine höhere Homologie zu Stämmen aus Gruppe 3 zu erkennen. Die Vermutung war, dass dies durch einen möglichen Verlust der *serU*-Insel zustande kam. Deswegen wurden die beiden inter-PAI-Regionen genauer untersucht. Hierfür wurden die Sequenzen der Regionen halbiert und erneut durch ein *Alignment* verglichen. Somit sollte das direkt umliegende Genom der HPI und der anderen beiden PAIs untersucht werden.

In Abbildung 37 ist das Genom direkt um die HPI zu sehen. In Abschnitt a) wird der Bereich ca. 500 bp *upstream* der HPI verglichen, in Abschnitt b) der Bereich ca. 6 kb *downstream*. In beiden Stammbäumen ist zu erkennen, dass sich jeweils drei phylogenetische Cluster gebildet haben, entsprechend den PAI-Gruppen. Auffällig ist, dass sich der Stamm ECOR65 jeweils in die PAI-Gruppe 3 eingliedert.

Die Abbildung 38 zeigt die Genomregionen um a) die *serU*-Insel (ca. 500 bp) und b) die *pks*-Insel (ca. 6 kb). Beim Vergleich der Sequenzen *upstream* der *pks*-Insel ist zu erkennen, dass die PAI-Gruppe 2b einen eigenen Cluster bildet, während sich die Gruppen 2a und 3 sowie der Stamm ECOR65 größtenteils durchmischen. Die umliegende Region der *serU*-Insel zeigt, dass die Sequenz aller Stämme identisch ist, nur die der Stämme ECOR65 und ECOR63 nicht. Allerdings besteht diese Abweichung aus jeweils nur einem SNP pro Stamm, was als nicht aussagekräftig eingestuft wird.

Das Ergebnis dieser Analysen und der spezifische *pks*-Insel-Subtyp lassen vermuten, dass der Stamm ECOR65 ursprünglich zur PAI-Gruppe 3 gehörte und möglicherweise die *serU*-Insel verloren hatte. Allerdings spricht die geringe Homologie zwischen den HPIs der PAI-Gruppe 3 und der von ECOR65 gegen diese Hypothese. Ein Vergleich mit den untersuchten ICE-Typ-HPIs zeigt keine Ähnlichkeit der ECOR65-HPI zu diesen.



a)



b)

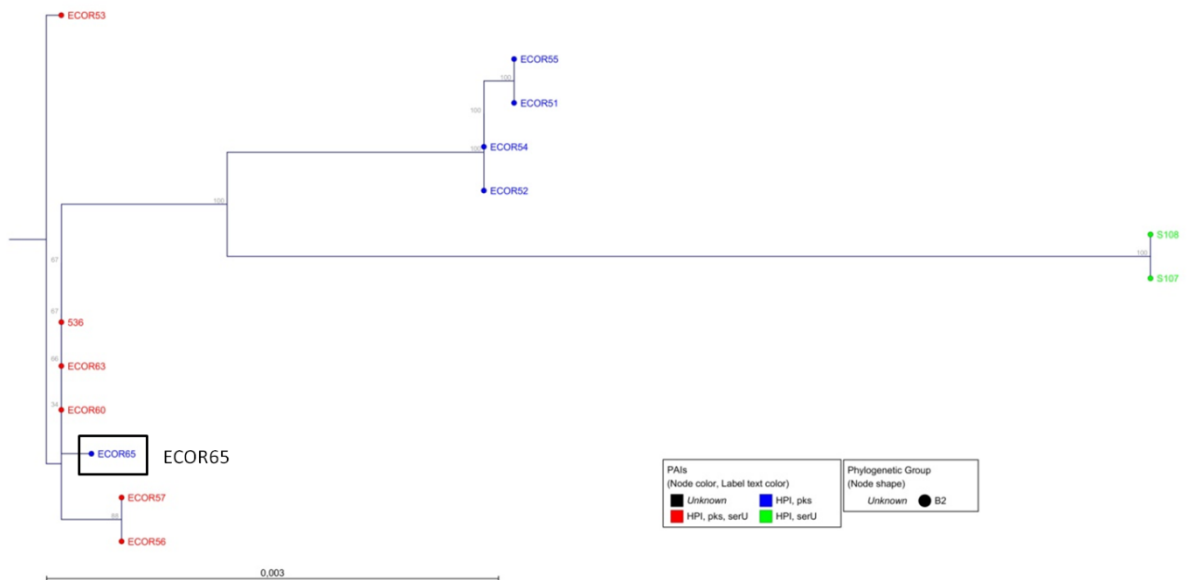
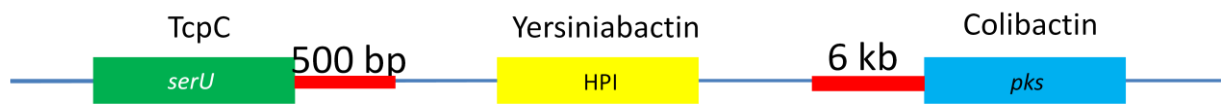
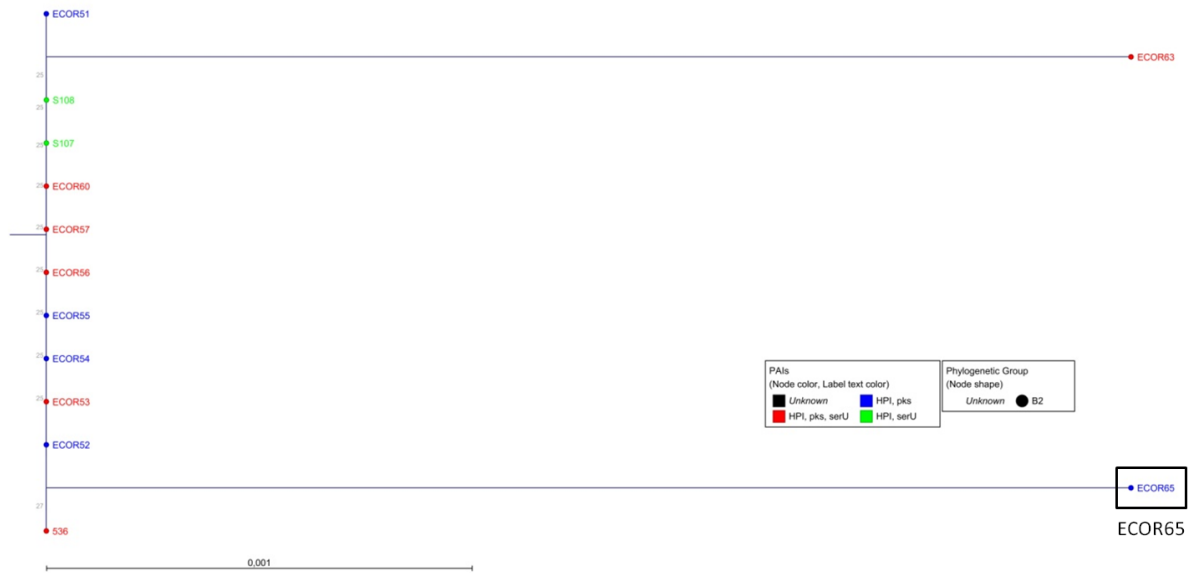


Abbildung 37: Untersuchung der um die HPI liegenden inter-PAI-Regionen. Der Stamm ECOR65 ist hervorgehoben. a) Vergleich der Sequenzen mit einer Größe von ca. 500 bp *upstream* der HPI. b) Vergleich der Sequenzen mit einer Größe von ca. 6 kb *downstream* der HPI.



a)



b)

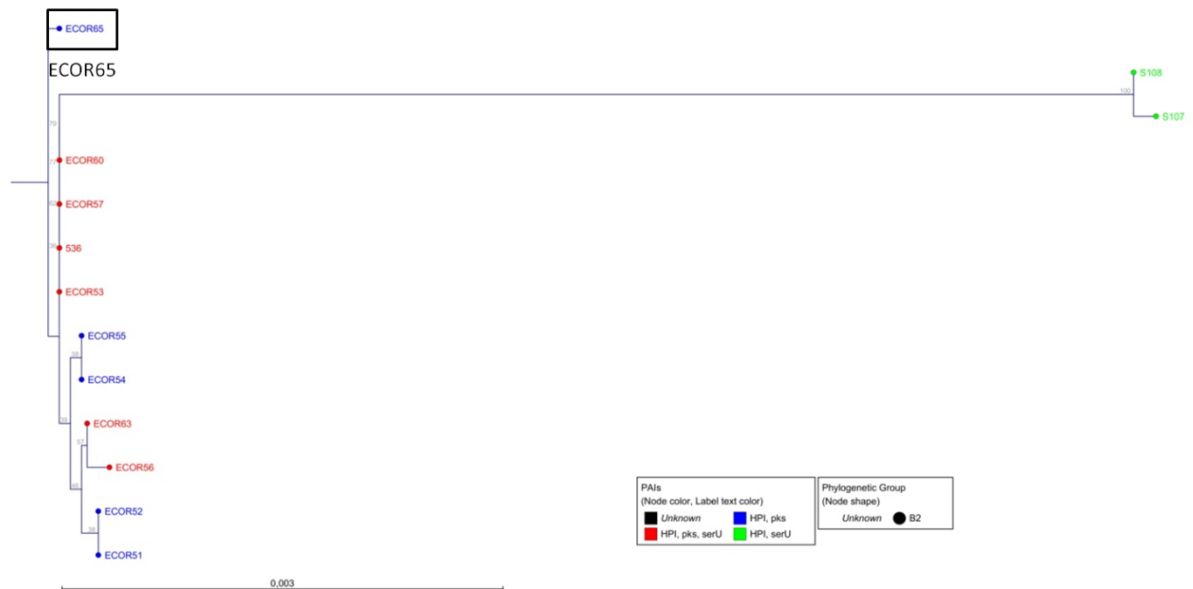


Abbildung 38: Untersuchung der Genomregionen direkt neben der *serU*- und der *pks*-Insel. Der Stamm ECOR65 ist hervorgehoben. a) Vergleich der ca. 500 bp langen Sequenz *downstream* der *serU*-Insel. b) Vergleich der ca. 6 kb langen Sequenz *upstream* der *pks*-Insel.

3.8. Untersuchung weiterer ECOR-Stämme

Bei der Betrachtung des HPI-Stammbaumes, der in Abbildung 22 zu sehen ist, ist zu erkennen, dass gewisse Stämme aus der PAI-Gruppe 1 weitere phylogenetische Cluster bilden. Die Stämme ECOR59 und ECOR66 bilden eine Gruppe, die zur HPI der PAI-Gruppe 3 eine erhöhte Homologie zeigt. Die Sequenz der Insel von Stamm ECOR64 reiht sich zu der HPI der PAI-Gruppe 2b ein. Die Stämme ECOR38 - 41 und ECOR46 bilden einen Cluster, der zwischen den PAI-Gruppen 2a und 2b liegt. Die PAIs der Stämme ECOR35 und ECOR36 haben die größte Übereinstimmung mit der des Stammes ECOR65. Da die HPIs der aufgeführten Stämme eine gewisse Homologie zu den anderen PAI-Gruppen zeigt, wurden diese ebenfalls auf den Verlust von PAIs überprüft. Dafür wurde dasselbe Verfahren angewandt wie für den Stamm ECOR65. Die Sequenzen der inter-PAI-Regionen wurden ebenfalls halbiert und als Stammbaum dargestellt. Die erstellten Bäume zeigen, dass sich die untersuchten Stämme größtenteils in eigene Cluster gruppieren und keine Zugehörigkeit zu anderen PAI-Gruppen erkennbar ist (Daten nicht dargestellt). Somit kann ein PAI-Verlust ausgeschlossen werden.

4. Transfer der immobilen PAIs

Um die Hypothese des *en bloc*-Transfers zu beweisen, sollte der Transfer von mehreren Inseln experimentell nachgestellt werden. Da die untersuchten PAIs immobil sind, wurde ein F'-Plasmid für die Übertragung gewählt. Es wurden zwei Konjugationen geplant, wofür zuerst die Donoren erstellt werden mussten. Es sollten die PAIs HPI und *pks*-Insel von dem Donor NU14 Cm F' auf den Rezipienten MG1655 übertragen werden und im Weiteren die PAIs HPI, *pks*- und *serU*-Insel von dem Donor 536 Cm F' auf den Rezipienten MG1655 Nal.

4.1. Erstellung der Donoren und Konjugation

Die beiden *E. coli*-Donor-Stämme NU14 Cm F' und 536 Cm F' wurden im Vorfeld unter den gleichen Bedingungen nach der Methode von Datsenko und Wanner (Datsenko und Wanner, 2000) mit einer Chloramphenicol-Resistenzkassette versehen. Es wurden für beide Konjugationen dieselben Bedingungen gewählt, welche nachfolgend erläutert werden (Abb. 39). Eine genaue Beschreibung ist unter "B.9.3.1. Konjugation von Pathogenitätsinseln mit einem F'-Plasmid" zu finden.

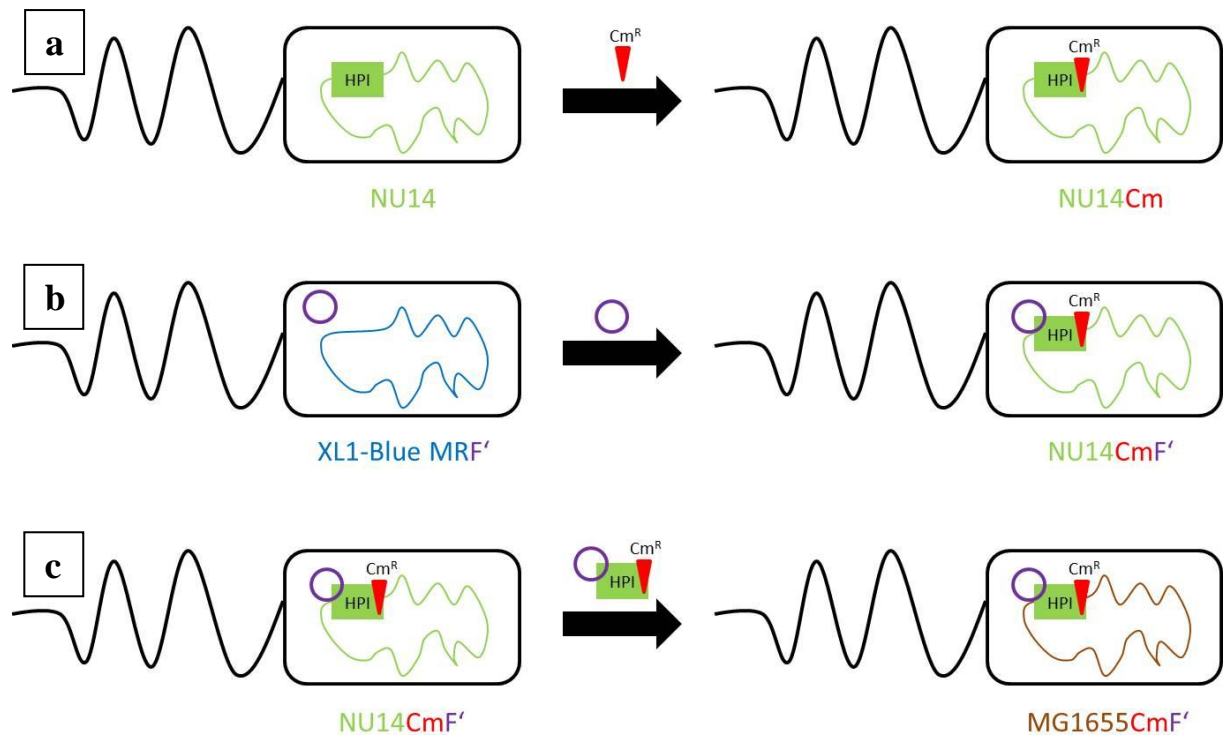


Abbildung 39: a) Insertion einer Chloramphenicol Resistenz Kassette in einen nicht codierenden Bereich der HPI *downstream* des Gens *fyuA*, b) Übertragung des F'-Plasmids vom Laborstamm *E. coli* XL1-Blue MRF' auf den zukünftigen Donor und c) Übertragung der PAIs mit Hilfe des F'-Plasmids auf den Rezipienten. Darstellung am Beispiel der Stämme NU14 und MG1655.

a) Anfangs musste die HPI der zukünftigen Donoren mit einer Chloramphenicol-Resistenzkassette nach der Methode von Datsenko und Wanner (Datsenko und Wanner, 2000) markiert werden, um die Übertragung der Inseln verfolgen zu können. Zum Erhalt dieser Resistenzkassette wurde mit dem Plasmid pKD3 eine PCR mit den Primern P1-*fyuA*-2104.for und P2-*fyuA*-2691.rev durchgeführt. Nach der Aufnahme des für die Insertion benötigten Plasmids pKD46 durch die Bakterienstämme wurde das PCR-Produkt in diese transformiert. Die Selektion der Chloramphenicol-resistenten Bakterien erfolgte durch das Ausplattieren auf LB-Agarplatten mit dem genannten Antibiotikum.

b) Da die PAIs nicht selbst-transferierbar sind, sollte als Hilfsmittel das F'-Plasmid des *E. coli*-Stammes XL1-Blue MRF' verwendet werden. Durch eine Konjugation wurde dieses Plasmid auf die ausgewählten Stämme übertragen, welche durch Chloramphenicol und Tetracyclin selektiert wurden. Die gewachsenen Klone wurden auf das Vorhandensein der HPI, auf deren phylogenetischen Gruppe und auf die Produktion von β -Hämolysin überprüft.

c) Für die Konjugation und den entsprechenden Transfer der PAIs wurden die Stämme NU14 Cm F' und 536 Cm F' mit den *E. coli*-Rezipienten MG1655 und MG1655 Nal vermischt. Mit Hilfe des F'-Plasmids konnten die mit Antibiotika-Resistenzkassetten-markierten PAIs von den Donoren auf die Rezipienten übertragen werden. Die auf LB-Agarplatten mit Chloramphenicol und Streptomycin bzw. Nalidixinsäure gewachsenen Kolonien wurden gezählt und überprüft (PAIs, phylogenetische Gruppe, β -Hämolysin). Die Konjugationseffizienz wurde berechnet als Verhältnis von Transkonjuganten zu Donoren, in Kolonie-bildenden Einheiten pro Milliliter (cfu/ml) (Beaber et al., 2004).

Um die Konjugationseffizienz für den reinen F'-Plasmid-Transfer, unabhängig von dem Inhalt der übertragenen DNA, zu berechnen, wurde die Konjugation von NU14 Cm F' und MG1655 unter den bereits beschriebenen Konjugationsbedingungen wiederholt. Es wurden Verdünnungsreihen von 10^{-1} bis 10^{-5} ausplattiert. Diesmal wurde zur Selektion Streptomycin und Tetracyclin verwendet, um die Rezipienten, die das Plasmid erhalten hatten, zu identifizieren. Insgesamt wurden 120 Transkonjuganten auf das Vorhandensein einer Chloramphenicol-Resistenz überprüft.

Zur Überprüfung, wie schnell die Transkonjuganten ihre Tetracyclin-Resistenz und damit auch das F'-Plasmid ohne Antibiotika-Druck verlieren, wurden 10 Tetracyclin-resistente Klone verwendet. Diese wurden sowohl auf LB-Agarplatten mit Chloramphenicol und auf Platten mit Chloramphenicol und Tetracyclin ausgestrichen und \approx 18 h inkubiert. Am nächsten Tag wurde das Wachstum überprüft und Bakterien von der Agarplatte mit Chloramphenicol erneut auf zwei neue Platten mit den erwähnten Antibiotika ausgestrichen. Dieses Verfahren wurde drei Tage lang wiederholt.

Des Weiteren wurde ein Tetracyclin-resistenter Transkonjugant (MG1655 HPI-Cm F') als Donor eingesetzt. Als Rezipient wurde der Stamm MG1655 Nal verwendet. Die beschriebenen Konjugationsbedingungen wurden eingehalten und die Konjugationseffizienz berechnet.

4.2. in vitro-Auswertung der Transkonjuganten und Konjugationseffizienz

Um die Echtheit der Transkonjuganten zu prüfen, wurden diese auf verschiedene Eigenschaften getestet. Die phylogenetische Gruppe wurde mittels PCR mit den Primern ChuA.1, ChuA.2, YjaA.1, YjaA.2, TspE4C2.1 und TspE4C2.2 definiert (Clermont et al., 2000). Das Vorhandensein der PAIs wurde ebenfalls per PCR festgestellt. Die verwendeten

Primer waren für die HPI *fyuA.1080.for* und *fyuA.1709.rev*, für die *pks*-Insel *clbA.IHE.for* und *clbA.IHE.rev*, für die *serU*-Insel *tcpC.for* und *tcpC.rev* und für die PAI III *iroN.28.for* und *iroN.2053.rev*. Anhand des Ausstreichens auf LB-Agarplatten mit entsprechenden Antibiotika wurden die Resistenzen analysiert. Durch das Auftragen auf Blut-Agarplatten wurde die Fähigkeit zur β -Hämolyse untersucht.

Der Transfer der PAIs HPI und *pks*-Insel von dem erstellten Donor NU14 Cm F' auf den Rezipienten MG1655 war erfolgreich und wurde durch die Selektion auf LB-Agar mit Chloramphenicol und Streptomycin nachgewiesen. Auf den Kontrollplatten (NU14 auf LB-Streptomycin und MG1655 auf LB-Chloramphenicol) waren keine Klone gewachsen. Es wurde eine Konjugationseffizienz von $2,94 \times 10^{-7}$ cfu/ml errechnet. Insgesamt 8% der Transkonjuganten erhielten keine *pks*-Insel. Die PAI III konnte in keinem Transkonjuganten nachgewiesen werden. Ungefähr 50% der erhaltenen Transkonjuganten hatten eine Tetracyclin-Resistenz. Es wurde untersucht, wie schnell diese Stämme die Resistenz und damit das F'-Plasmid verlieren. Hierfür wurden 10 Tetracyclin-resistente Transkonjuganten auf LB-Agarplatten mit Chloramphenicol und Streptomycin aber ohne Tetracyclin inkubiert, um keinen Druck zu erzeugen, das Plasmid zu behalten. Nach der ersten Passage hatten 20% ihre Resistenz verloren und nach der zweiten bereits 60%. Nach dreimaliger Inkubation auf LB-Agarplatten ohne entsprechenden Antibiotika-Druck waren 90% von diesen nicht mehr Tetracyclin-resistent. Um die Weitergabe des F'-Plasmids ohne Berücksichtigung auf die HPI zu ermitteln, wurde der Konjugationsansatz auf LB-Agar mit Tetracyclin und Streptomycin ausplattiert. Die hierfür berechnete Konjugationseffizienz beträgt $5,24 \times 10^{-4}$ cfu/ml. Von 120 untersuchten Stämmen besitzt keiner eine Chloramphenicol-Resistenz und somit auch keine HPI. Des Weiteren wurde der Transkonjugant MG1655 HPI-Cm F' als Donor eingesetzt. Als Rezipient diente der Stamm MG1655 Nal. Die Konjugationseffizienz beträgt in diesem Fall $4,14 \times 10^{-5}$ cfu/ml und wurde durch die Selektion auf LB-Agar mit Chloramphenicol und Nalidixin nachgewiesen.

Auch die Konjugation der drei PAIs HPI, *pks*- und *serU*-Insel von dem erstellten Donor 536 Cm F' in den Rezipienten MG1655 Nal war erfolgreich und wurde durch die Selektion auf LB-Agar mit Chloramphenicol und Nalidixin bestätigt. Auf den Kontroll-Platten (536 auf LB-Nalidixin und MG1655 Nal auf LB-Chloramphenicol) war kein Wachstum zu erkennen. Die Konjugationseffizienz beträgt $3,85 \times 10^{-7}$ cfu/ml. Insgesamt 10% der Transkonjuganten erhielten keine *pks*-Insel. Die PAI III wurde nicht auf die Transkonjuganten übertragen.

4.3. in silico-Auswertung der Transkonjuganten

Es wurden die Genome von insgesamt neun Transkonjuganten aus unabhängigen Konjugationen sequenziert, um das Ausmaß des Transfers zu begutachten (Tab. 25). Hiervon resultierten fünf aus dem Donor NU14 Cm F' (Klon 1 - 5) und vier aus dem Donor 536 Cm F' (Klon 6 - 9). Von beiden Transferversuchen wurde jeweils ein Stamm ohne *pks*-Insel untersucht. Resultierend aus dem Donor NU14 Cm F' wurde zusätzlich ein Tetracyclin-resistenter Stamm (Klon 1 TetR) gewählt, der solange passagiert wurde, bis er seine Resistenz verloren hatte (Klon 1). Beide Varianten wurden in die Sequenzierung mit eingeschlossen. Bei allen Transkonjuganten wurden vollständige PAIs gefunden, es wurden also keine Bruchstücke der Inseln übertragen. Es wurde nur ein DNA-Block transferiert, nicht verschiedene einzelne Stücke. Auch *in silico* konnte kein PAI III-Nachweis erbracht werden.

Durch einen Vergleich mit dem Donor und dem Rezipienten wurde für jeden Transkonjuganten die Menge der übertragenen DNA bestimmt. Die Ergebnisse sind in Tabelle 25 zu sehen. Die Menge übertragener DNA variiert zwischen ca. 62,5 kb und ca. 470 kb und ist von dem Donor unabhängig. Bei allen untersuchten Transkonjuganten, die eine *pks*-Insel erhielten, bricht das *Contig* 12.250 bp \pm 1 bp *downstream* dieser Insel ab. Bei drei dieser Stämme endet tatsächlich die Donor-DNA in dieser Region, das nächste passende *Contig* wurde als Rezipienten-DNA identifiziert. Bei den anderen Stämmen endet die Donor-DNA erst weitaus später. Alle Rekombinationsorte der Donoren-DNA wurden in dem Stamm MG1655 auf mobile Elemente überprüft, es konnten aber keine identifiziert werden.

Tabelle 25: Größe übertragener DNA-Stücke bei der Nachstellung des *en bloc*-Transfers. Die Transkonjuganten entstanden entweder aus dem Donor NU14 (Klone 1 - 5) oder 536 (Klone 6 - 9). Gezeigt wird die Menge an DNA *upstream* der HPI und *downstream* der *pks*-Insel. Bei Stämmen, die keine *pks*-Insel erhalten hatten, wurde die DNA-Menge *downstream* der HPI angegeben. Es wurde die gesamte übertragene DNA-Menge aufgelistet.

NU14 x MG1655	<i>upstream</i> HPI (bp)	<i>downstream</i> pks (bp)	insgesamt (bp)
Klon 1	18.344 – 18.804	106.404 – 106.725	223.368 – 224.149
Klon 1 TetR	18.344 – 18.804	106.404 – 106.725	223.368 – 224.149
Klon 2	61.676 – 61.753	161.659 – 161.700	321.955 – 322.073
Klon 3	309.303 – 309.338	12,2 kb < x < 13,1 kb	419.923 – 421.058
Klon 4	20.512 – 20.649	12,2 kb < x < 13 kb	131.132 – 132.269
Klon 5 (pks-)	164.328 – 164.367	2.484 – 2.511 (HPI)	198.334 – 198.400
536 x MG1655 Nal	<i>upstream</i> HPI (bp)	<i>downstream</i> pks (bp)	insgesamt (bp)
Klon 6	114.391 – 114.516	12,2 kb < x < 12,9 kb	225.011 – 226.036
Klon 7	181.494 – 181.834	190.075 – 190.137	470.189 – 470.591
Klon 8	124.694 – 124.864	124.853 – 124.924	348.167 – 348.408
Klon 9 (pks-)	28.268 – 28.376	2.706 – 2.813 (HPI)	62.496 – 62.711

Die Stelle, an der die *Contigs* ca. 12 kb *downstream* der *pks*-Insel enden, wurde weiter untersucht, um eine mögliche Rekombinationsstelle des F'-Plasmids zu identifizieren, die z.B. durch Insertionssequenzen (IS-Elemente) begünstigt wird.

Eine BLAST-Analyse der letzten bekannten 70 bp zeigte, dass sich diese Sequenz in den Stämmen NU14 und 536 am Ende der Glykosyl-Transferase *YaiP* befindet. An dieser Position besitzt der Stamm MG1655 keine Glykosyl-Transferase im Genom. Diese 70 bp sind bei den Stämmen NU14 und 536 nochmals innerhalb eines Transferase-Gens, das dieselbe Länge wie *yaiP* hat, *downstream* der PAI III zu finden (Abb. 40). An dieser Stelle besitzt der Stamm MG1655 das Glykosyl-Transferase-Gen *yaiP*. In diesem Stamm sind in den untersuchten 70 bp zwei SNPs zu finden. *YaiP* überlappt mit 47 Codons den N-Terminus von *YaiX*, was ungewöhnlich ist (Zhou und Rudd, 2013). Die letzten bekannten 70 bp sind 29 bp vor dieser Überlappung zu finden. In das Pseudogen *yaiX* des Stammes MG1655 ist ein IS-Element (IS2a) integriert, allerdings besitzt der Stamm 536 ein intaktes *yaiX* (Zhou und Rudd, 2013). Dasselbe IS-Element konnte auch *in silico* auf dem F'-Plasmid nachgewiesen werden. Das Gen *yaiX* gehört zu der *hexapeptide transferase superfamily* und besitzt sechs Hexapeptid Wiederholungen (*repeats*) (Zhou und Rudd, 2013).

Die Stellen, an denen die Transferasen im Genom zu finden sind, sind anhand des Stammes 536 in Abbildung 40 zu sehen. Es konnte innerhalb der ECOR-Sammlung keine Korrelation zwischen dem Vorhandensein der Glykosyl-Transferasen und der HPI gefunden werden.

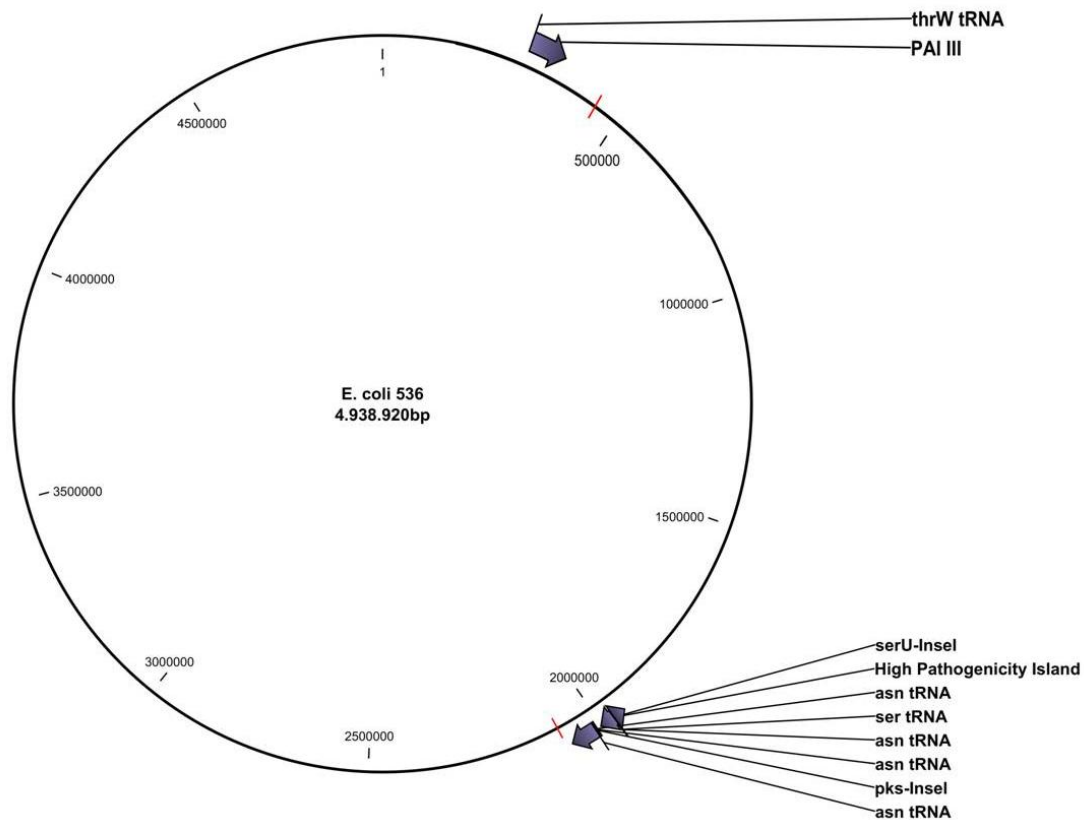


Abbildung 40: Genomkarte des *E. coli*-Stammes 536. Die beiden untersuchten Transferasen sind durch rote Striche gekennzeichnet. Die PAIs sind als Pfeile dargestellt. Die entsprechenden tRNAs sind angegeben.

Um zu überprüfen, ob ein Grund für den Abbruch der *Contigs* 12 kb *downstream* der *pks*-Insel gefunden werden kann, sollte der fehlende Abschnitt sequenziert und auf mögliche Insertionen von IS-Elementen untersucht werden. Mit den Primern 12kb.pks.for und 12kb.pks.rev wurde eine PCR mit allen Transkonjuganten sowie den Donoren und dem Rezipient MG1655 NaI durchgeführt. In Abbildung 41 ist das Gelbild zu sehen. Es sind meist zwei PCR-Banden zu erkennen, eine mit der Größe von ca. 500 bp und eine von ca. 2 kb. Die PCR-Produkte wurden aufgereinigt und mit demselben Primer-Paar sequenziert.

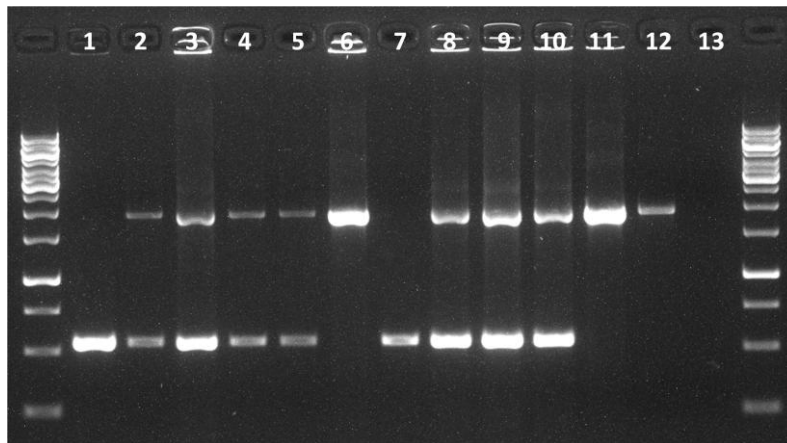


Abbildung 41: PCR mit den Primern 12kb.pks.for und 12kb.pks.rev. Stämme von links nach rechts: 1) NU14; 2) K1; 3) K2; 4) K3; 5) K4; 6) K5 (pks-); 7) 536; 8) K6; 9) K7; 10) K8; 11) K9 (pks-); 12) MG1655 Nal; 13) Leerkontrolle. Marker: GeneRuler™ 1 kb.

Es stellte sich heraus, dass die kleinere Bande zu den Stämmen 536 und NU14 passt. Es sind keine weiteren DNA-Insertionen an dieser Stelle zu finden, die für den Abbruch der Sequenzierung verantwortlich sein könnten. Auch F'-Plasmid-DNA ist hier nicht vorhanden. Die größere Bande ist ein Teilstück der bereits erwähnten Transferase des Stammes MG1655, die ein IS-Element enthält. Bei genauer Betrachtung des Gelbildes ist zu erkennen, dass die Donoren NU14 (1) und 536 (7) nur die kleinere Bande aufzeigen, während der Rezipient MG1655 Nal (12) nur die größere Bande aufweist. Die Transkonjuganten mit erhaltener *pks*-Insel (2 - 5; 8 - 10) besitzen die Transferasen von Donor und Rezipient, die ohne *pks*-Insel (6 und 11) nur die des Rezipienten. Die *in silico*-Untersuchung zeigte, dass die Sequenz der kleineren Bande eine Länge von 544 bp hat. Der Primer 12kb.pks.for hatte in dem Gen *yaiP* gebunden und der Primer 12kb.pks.rev in *yaiX*. In der Sequenz der größeren Bande mit einer Länge von 1.880 bp hatten die Primer in denselben Genen gebunden. Bei dem Stamm MG1655 sind in dem Gen *yaiX* die Gene *insD1* und *insC1* integriert, die zusammen das IS-Element mit der Länge von 1.331 bp bilden. Das Gen *yaiP* von MG1655 ist homologer zu dem Gen der Glykosyl-Transferase *downstream* der PAI III von *E. coli* 536 (97,33%; 32 SNPs) als zu dem Gen *yaiP* (94,65%; 64 SNPs) von diesem Stamm. Die Sequenzen der beiden zuletzt genannten Gene haben eine Homologie von 94,74% mit 63 SNPs.

Es wurden die Transkonjuganten Klon 1 und Klon 1 TetR miteinander verglichen, um den Verlauf der DNA-Integration in das Chromosom zu untersuchen. Da die Tetracyclin-Resistenz auf dem F'-Plasmid lokalisiert ist, war es möglich, dass die Inseln noch episomal vorlagen und die Rezipienten-DNA nicht ersetzt hatten. Laut Sequenzierungsdaten ist in

beiden Stämmen die Donoren-DNA mit den PAIs in das Rezipienten-Genom integriert. Bei dem Stamm Klon 1 TetR konnte keine integrierte F'-Plasmid-DNA festgestellt werden. Die Untersuchung der Rohdaten zeigt, dass in diesem Stamm die DNA des Donors die des Rezipienten ersetzt hat. Es wurden keine Überreste der ursprünglichen Rezipienten-DNA gefunden, auch nicht episomal auf dem F'-Plasmid. Die Menge an übertragener DNA ist ebenfalls in beiden Stämmen identisch (Tab. 25).

5. Phylogenetische Auswertung der mobilen PAIs

Als Beispiel für die Phylogenie von mobilen PAIs wurde die ICE-Typ-HPI gewählt. Bestimmte Regionen von dieser wurden zusätzlich mit den bereits analysierten *E. coli*-PAIs verglichen. Für die Generierung der Stammbäume wurden Sequenzen von Bakterienstämmen der Familie *Enterobacteriaceae* verwendet. Hierfür dienten Genome, die sowohl durch die eigene Sequenzierung erstellt wurden, aber auch bereits veröffentlichte. Die im Rahmen dieser Arbeit erstellten Sequenzen waren von den Stämmen *K. pneumoniae* HU86, HU145, HU245, B-2495 und *E. coli* ECOR31. Die vom *National Center for Biotechnology Information* (NCBI) öffentlich verfügbaren Genomsequenzen waren von den Stämmen *K. pneumoniae* 1084 (NCBI-Zugangsnummer CP003785), *Citrobacter koseri* ATCC BAA-895 (NCBI-Zugangsnummer NC_009792) und *E. coli* ED1a (NCBI-Zugangsnummer NC_011745).

5.1. Konservierte Kern-Region

Der Stammbaum in Abbildung 42 vergleicht die Kern-Region, also den konservierten Bereich der mobilen HPI, der für die Synthese und den Transport des Siderophors Yersiniabaktin zuständig ist (Schubert et al., 2004a). Es wurde auch eine statistische Analyse angefertigt, die in Tabelle 26 zu sehen ist. Es ist zu erkennen, dass die Sequenzen dieser Region innerhalb der Art *E. coli* einen eigenen Cluster mit einer Homologie von 99,30% (221 SNPs) bilden. Dies ist auch der Fall innerhalb der Art *K. pneumoniae*, allerdings nur für ICEs, die die Region X enthalten, aber keine *pks*-Insel innerhalb der Region III beinhalten. Diese Gruppe hat die höchste Homologie mit 99,96% (13,3 SNPs). Die Sequenz des Stammes HU245 ordnet sich nicht zu denen der anderen *K. pneumoniae*-Stämme ein. Die Stämme *C. koseri* 895 und *K. pneumoniae* 1084, die die zusätzliche *pks*-Insel innerhalb des ICE enthalten, sind weniger

homolog zueinander, bilden aber einen eigenen Cluster (98,36%; 520 SNPs). Innerhalb der Art *K. pneumoniae* ist eine Homologie von 99,49% mit durchschnittlich 161,7 SNPs zu finden. Die Sequenzen der konservierten Kern-Region aller untersuchten Stämme mit einer ICE-Typ-HPI haben eine Homologie von 99,22% (247,2 SNPs). Im Vergleich dazu ist die Homologie der kürzeren *E. coli*-HPI 99,53% (149,7 SNPs).

Tabelle 26: Statistische Auswertung der Kern-Region innerhalb der untersuchten ICEs. Berechnet wurde die Homologie und die durchschnittliche Anzahl an SNPs.

untersuchte Kern-Region	Homologie	SNPs
innerhalb <i>E. coli</i> (n=1)	99,30%	221
innerhalb <i>K. pneumoniae</i> (n=10)	99,49% (99,02% - 100%)	161,7 (1 - 310)
RegX-positive <i>K. pneumoniae</i> (n=3)	99,96% (99,94% - 100%)	13,3 (1 - 20)
ICE mit <i>pks</i> -Insel in Region III (n=1)	98,36%	520
alle Sequenzen (n=28)	99,22% (98,36% - 100%)	247,2 (1 - 520)

Um einen Vergleich zwischen der ICE-Typ-HPI und der kurzen HPI zu ziehen, wurde ein Stammbaum erstellt, der die konservierten Kernregionen aller untersuchten Bakterienstämme darstellt (Abb. 43). In diesem ist zu erkennen, dass sich die meisten der neu eingefügten ICE-Typ-Sequenzen zu der des Stammes ECOR31 anordnen, der ebenfalls eine ICE-Typ-HPI besitzt. Die Sequenz des Stammes *K. pneumoniae* HU245, dessen HPI weder eine Region X noch eine *pks*-Insel beinhaltet, bildet einen Cluster mit dem Stamm ECOR72. Es wurde bereits vermutet, dass der Stamm ECOR72 eine intermediäre Form zwischen der ICE-Typ-HPI und der kurzen HPI darstellt (Schubert et al., 2009). Die HPI-Sequenz des Stammes ECOR65, der laut Definition in die PAI-Gruppe 2a einzuordnen ist, zeigt in diesem Stammbaum keine nähere Verwandtschaft zu den untersuchten Sequenzen der ICE-Typ-HPIs.

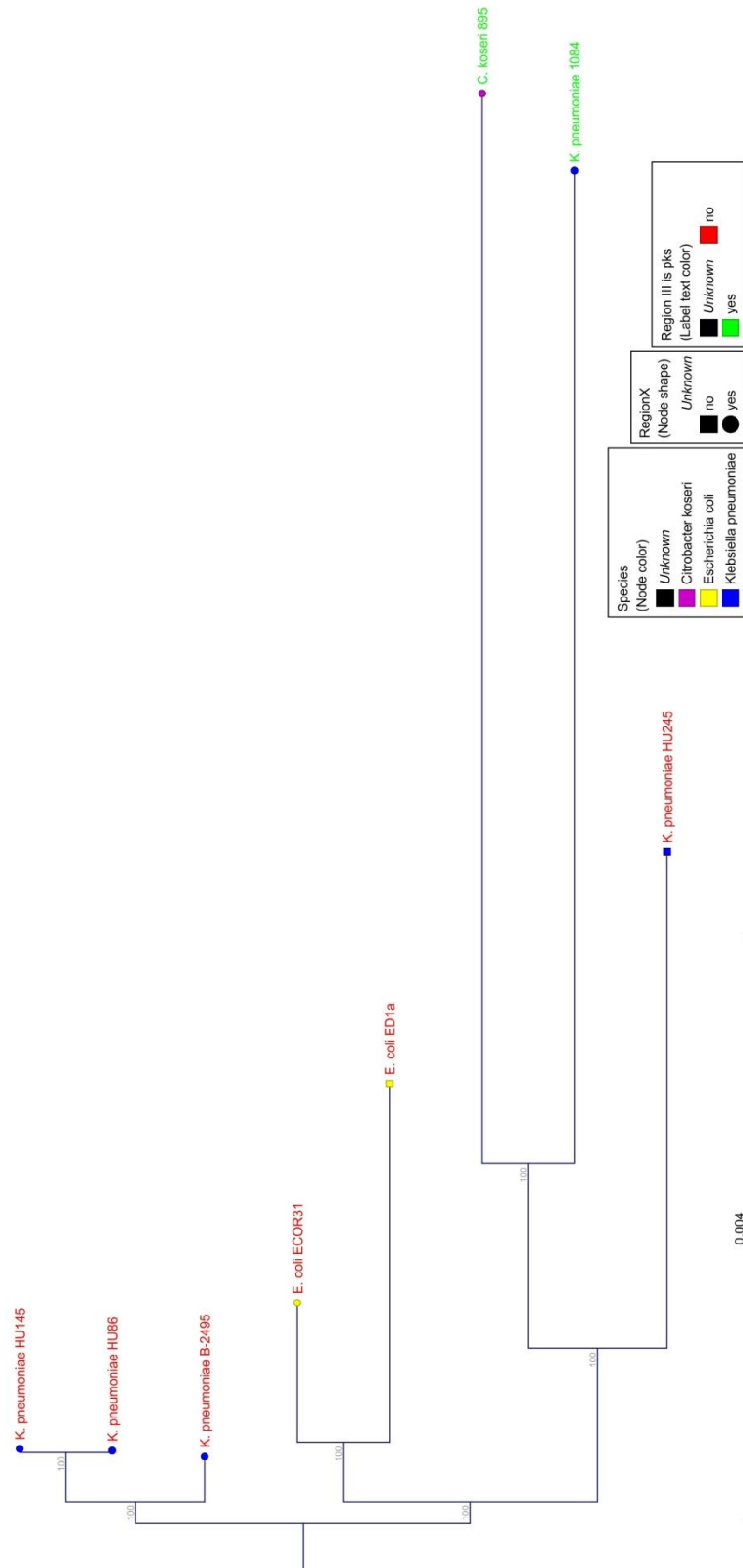


Abbildung 42: Stammbaum der konservierten Kern-Region der ICE-Typ-HPI. In der Legende sind die Bakterienart und das Vorhandensein der Region X und der *pks*-Insel in der HPI aufgelistet. Der Maßstab zeigt die Anzahl an SNPs pro Base.

5.2. Region X

Die Region X hat eine hohe Homologie zu dem Virulenz-Plasmid pLVPK von *K. pneumoniae* und ist bei manchen ICE-Typ-HPIs zwischen der Integrase *intB* und der *asn* tRNA zu finden (Preisach, 2008). Diese Region wurde bei vorliegenden Isolaten mittels PCR und bei öffentlich verfügbaren Sequenzen *in silico* ermittelt. In beiden Fällen wurde dies mit den Primern RegX-521.for und RegX-358.rev vollzogen. Die Region X ist in den *K. pneumoniae*-Stämmen B-2495, HU86, HU145 und 1084 sowie in *E. coli* ECOR31 und *C. koseri* ATCC BAA-895 präsent.

Anhand des Stammbaums in Abbildung 44 wird die Region X innerhalb der untersuchten ICEs verglichen und die Sequenzen in Tabelle 27 statistisch ausgewertet. Die Stämme mit einer *pks*-Insel in der Region III bilden einen Cluster mit einer Homologie von 99,87% und 43 SNPs. Die Sequenzen der Region X der *K. pneumoniae*-Stämme HU86 und HU145 sind zu 100% identisch. Insgesamt ist innerhalb dieser Art eine Homologie von 99,92% mit durchschnittlichen 25,7 SNPs zu verzeichnen. Beim Vergleich aller Sequenzen ist eine hohe Homologie von 99,88% mit durchschnittlich 39,4 SNPs zu sehen, was auch deutlich am Maßstab erkennbar ist.

Tabelle 27: Statistische Auswertung der Region X innerhalb der untersuchten ICEs. Berechnet wurde die Homologie und die durchschnittliche Anzahl an SNPs.

untersuchte Region X	Homologie	SNPs
innerhalb <i>K. pneumoniae</i> (n=3)	99,92% (99,83% - 100%)	25,7 (0 - 56)
ICE mit <i>pks</i> -Insel in Region III (n=1)	99,87%	43
alle Sequenzen (n=15)	99,88% (99,75% - 100%)	39,4 (0 - 83)

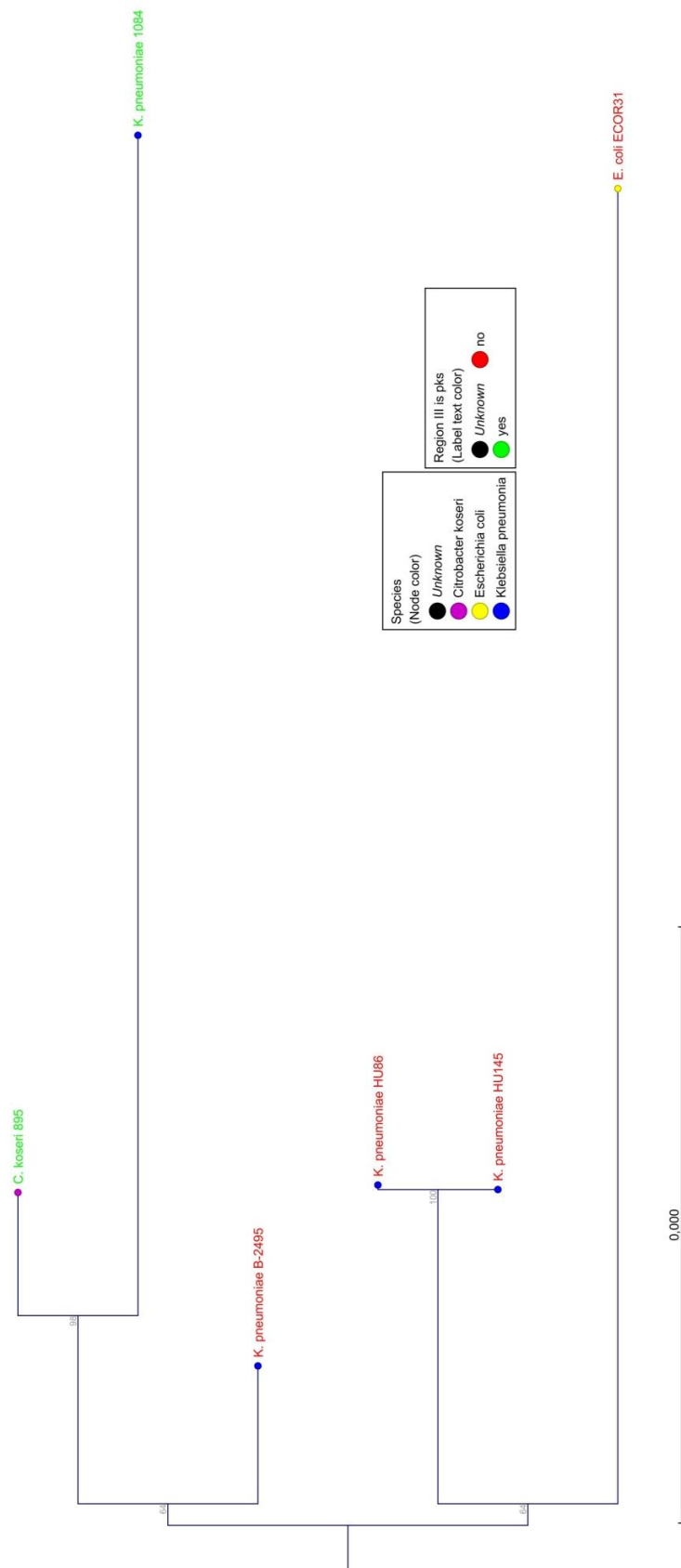


Abbildung 44: Stammbaum der Region X aller positiven Stämme. Die Bakterienart und das Vorhandensein der *pks*-Insel innerhalb der Region III der HPI sind in der Legende zu sehen. Der Maßstab beschreibt die Anzahl an SNPs pro Base.

5.3. Region III: *pks*-Insel

Die *pks*-Insel, die in der Region III des ICE von *C. koseri* ATCC BAA-895 und *K. pneumoniae* 1084 zu finden ist, ist nicht vollständig. Es fehlen bei beiden Inseln die anfänglichen 1.610 bp, einschließlich der Integrase *intB*. Reste einer ursprünglichen *asn* tRNA sind ebenfalls nicht vorhanden. Um zu untersuchen, wie homolog diese PAIs im Vergleich zu denen der *E. coli*-Stämme sind, wurden alle Sequenzen auf die Länge der ICE-Typ-*pks*-Insel gekürzt. Es wurden ein *Alignment* und ein Stammbaum erstellt und statistisch ausgewertet (Abb. 45; Tab. 28). Der phylogenetische Baum in Abbildung 45 zeigt, dass die ICE-*pks*-Inseln sehr homolog untereinander (99,91%; 49 SNPs), aber auch zu den anderen untersuchten PAIs der Art *E. coli* (99,89%; 58,1 SNPs) sind. Die Sequenzen der ICE-*pks*-Inseln bilden trotz hoher Homologie zu den *E. coli*-Inseln einen eigenen Cluster. Verglichen mit den PAI-Gruppen 2a und 3 wurde eine ähnliche Homologie von 99,90% (53,8 SNPs) und 99,89% (59,8 SNPs) errechnet. Die Übereinstimmung mit dem Stamm ECOR65 ist mit 99,88% und 64,5 SNPs im Vergleich zu den anderen Stämmen am niedrigsten.

Tabelle 28: Statistischer Vergleich der *pks*-Inseln von *C. koseri* 895 und *K. pneumoniae* 1084, den PAI-Gruppen und dem Stamm ECOR65. Aufgelistet wurde die Homologie und die durchschnittliche Anzahl an SNPs.

<i>pks</i> -Insel	Homologie	SNPs
Vergleich <i>C. koseri</i> mit <i>K. pneumoniae</i> (n=1)	99,91%	49
Vergleich ICE- <i>pks</i> mit <i>E. coli-pks</i> (n=22)	99,89% (99,82% - 99,95%)	58,1 (28 - 95)
Vergleich ICE- <i>pks</i> mit PAI-Gruppe 2a (n=8)	99,90% (99,85% - 99,95%)	53,8 (28 - 78)
Vergleich ICE- <i>pks</i> mit PAI-Gruppe 3 (n=12)	99,89% (99,82% - 99,95%)	59,8 (28 - 95)
Vergleich ICE- <i>pks</i> mit ECOR65 (n=2)	99,88% (99,84% - 99,92%)	64,5 (43 - 86)

Eine MLST mit den Genen *clbA*, *clbE*, *clbI*, *clbL*, *clbO* und *clbQ* (gesamte Sequenzlänge: 8.664 bp) ist möglich, da die Stämme *C. koseri* 895 und *K. pneumoniae* 1084 auch in den codierenden Genen der *pks*-Insel Unterschiede aufweisen. Der Stammbaum in Abbildung 46 zeigt außerdem, dass die Gene der *E. coli*-Stämme komplett identisch sind. Anhand der statistischen Auswertung kann man sehen, dass die Abweichungen in den sechs Genen von *C. koseri* 895 und *K. pneumoniae* 1084 jeweils nur auf zwei SNPs basieren und diese zu 99,98% homolog sind. Bei der Sequenz von *C. koseri* 895 sind beide SNPs in dem Gen *clbI* lokalisiert. Die SNPs in der Sequenz von *K. pneumoniae* 1084 sind in den Genen *clbE* und *clbI* zu finden, wobei im letzteren die Position des SNP derselben im *C. koseri*-Gen entspricht.

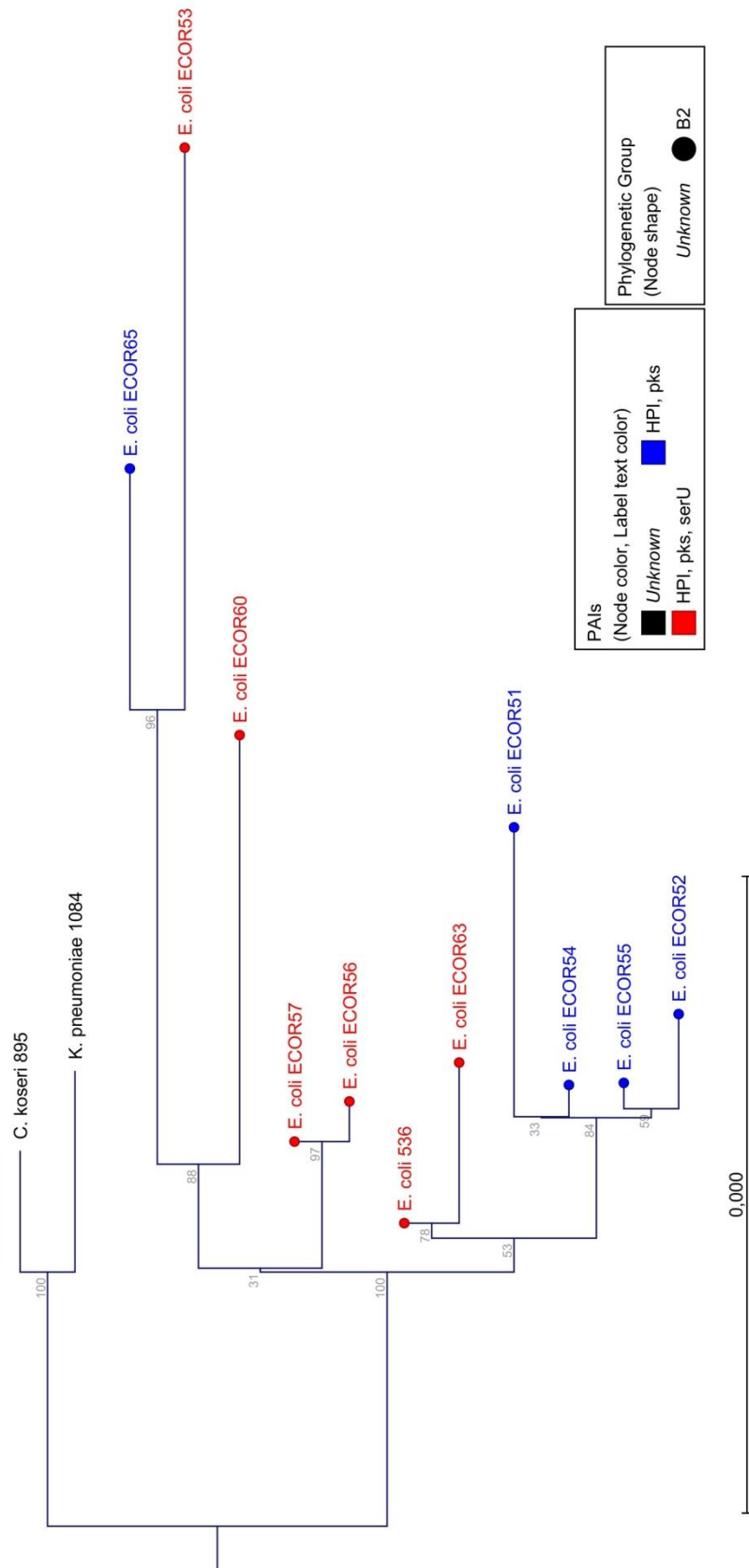


Abbildung 45: Stammbaum der *pks*-Insel. Die Sequenzen wurden auf die Länge der *pks*-Insel der Stämme *C. koseri* 895 und *K. pneumoniae* 1084 gekürzt. In der Legende sind für die *E. coli*-Stämme die PAI-Gruppen und die phylogenetische Gruppe zu sehen. Der Maßstab beschreibt die Anzahl an SNPs pro Base.

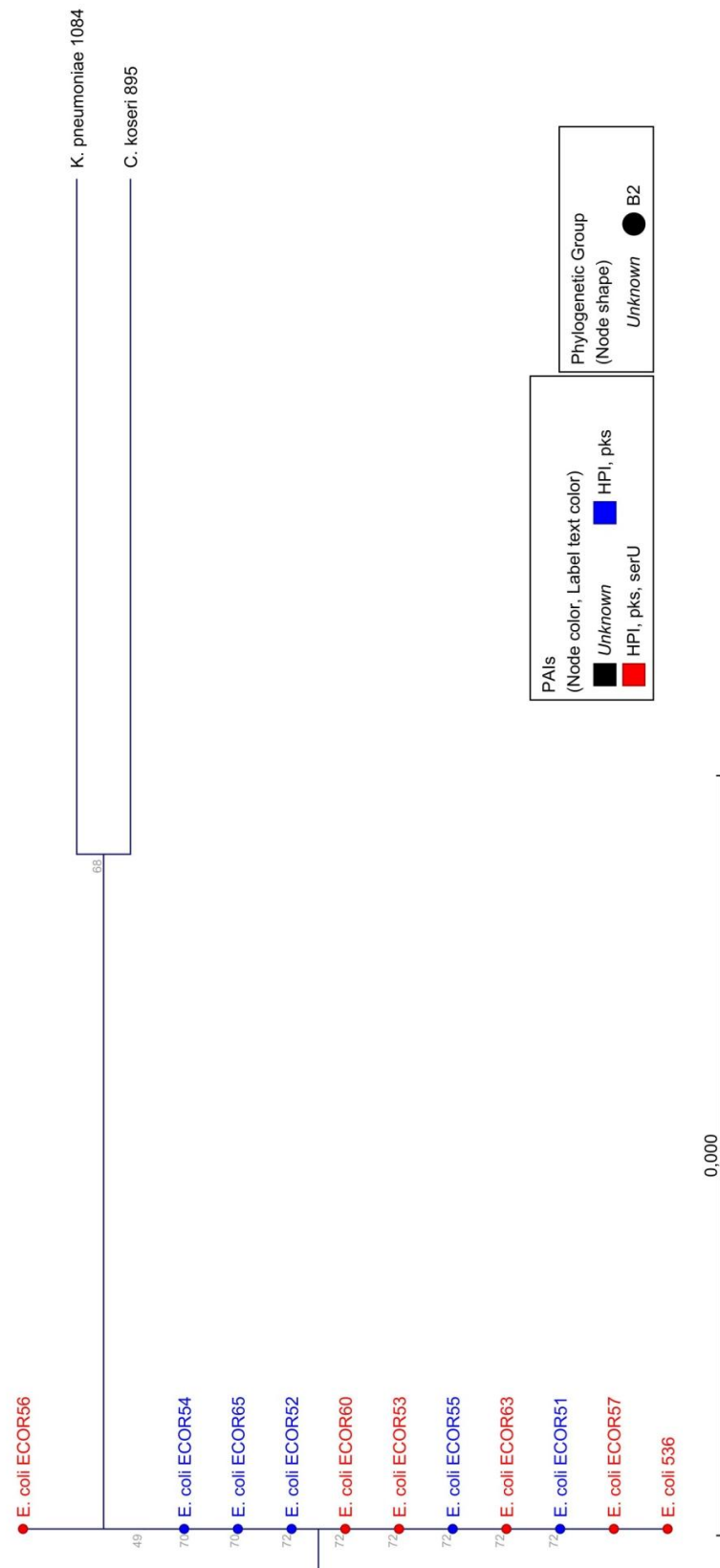


Abbildung 46: MLST-Stammbaum der *pks*-Insel. Verglichen wurden die sechs Gene *clbA*, *clbE*, *clbI*, *clbL*, *clbO* und *clbQ* (gesamte Sequenzlänge: 8.664 bp) zwischen den *E. coli*-Stämmen und den Stämmen *C. koseri* 895 und *K. pneumoniae* 1084. Der Maßstab gibt die Anzahl an SNPs pro Base an.

6. Transfer der mobilen PAI

Das Ziel dieser Versuche war, die ICE-Typ-HPI über die Gattungsgrenze hinaus zu übertragen. Um den Transfer zu verfolgen, mussten erst geeignete Donoren erstellt werden. Diese Stämme sind von den Arten *E. coli* und *K. pneumoniae*. Die Rezipienten, die die PAI erhalten sollten, sind *Y. enterocolitica*- und *E. coli*-Stämme.

6.1. Konstruktion der Donoren und Konjugation

Die Konstruktion der Donoren erfolgte ebenfalls nach der Methode von Datsenko und Wanner (Datsenko und Wanner, 2000) und ist detailliert unter dem Punkt "B.9.3.2. Konjugation von aktiv transferierbaren Pathogenitätsinseln" beschrieben. Es wurden PCRs mit den Primern P1-fyuA-2104.for und P2-fyuA-2691.rev durchgeführt, um die Antibiotika-Resistenzkassetten zu amplifizieren. Für Chloramphenicol wurde als DNA-Vorlage das Plasmid pKD3, für Kanamycin das Plasmid pKD4 verwendet. Nach der Aufnahme des für die Insertion benötigten Plasmids pKD46 durch die Bakterienstämme wurde das PCR-Produkt in diese transformiert. Die Stämme HU86 und ED1a erhielten eine Chloramphenicol-Resistenzkassette. Da der Stamm EDF72 bereits Chloramphenicol-resistent ist, erhielt dieser eine Kanamycin-Resistenzkassette.

Es wurden die unterschiedlichsten Konjugationsbedingungen ausgetestet, um eine möglichst hohe Übertragungsrate zu erzielen (siehe "B.9.3.2. Konjugation von aktiv transferierbaren Pathogenitätsinseln"). Für das Volumen der verwendeten Kulturen eignete sich am besten ein größerer Ansatz von 30 ml, sowohl für den Donor als auch für den Rezipienten. Als effektivste Wachstumsphase ergab sich die späte exponentielle Phase bei einer OD₆₀₀ von ca. 0,8, ebenfalls für Donor und Rezipient. Die Temperaturen für die Konjugationen erwiesen sich für *E. coli* bei 37°C am geeignetsten, für *Y. enterocolitica* bei 30°C. Als verwendetes Nährmedium bewährte sich eine Blut-Agarplatte. Ein Nitrocellulose-Filter war unumgänglich für einen erfolgreichen Transfer. Das Auslösen der SOS-Antwort im Donor erbrachte keinen Erfolg.

Die Konjugation war nur erfolgreich bei der Verwendung des Region X-positiven Donors HU86 Cm, sowohl in den Stamm *E. coli* MG1655 als auch in den Stamm *Y. enterocolitica* MRS40. Zusammengefasst war ein Transfer nur unter folgenden Bedingungen möglich:

Es wurden üN-Kulturen der Donor- und Rezipienten Stämme verwendet. Von den Kulturen wurde jeweils 1 ml in 29 ml frisches LB-Medium mit entsprechendem Antibiotika gegeben

und bei 37°C und 200 rpm bis zu einer OD₆₀₀ von ca. 0,8 inkubiert. Anschließend wurden die einzelnen Stämme bei 4.300 rpm für 15 min zentrifugiert und in 30 ml PBS durch Schwenken resuspendiert. Nach erneuter Zentrifugation war der Waschschrift abgeschlossen. Die Pellets wurden final in 300 µl PBS mit einer abgeschnittenen Pipetten-Spitze resuspendiert, um die Scherkräfte und damit den Pili-Abbruch zu reduzieren. Der Donor und der Rezipient wurden vermischt, und auf einen auf einer Blut-Agarplatte liegenden Nitrocellulose-Filter gegeben. Die Platte wurde üN in einem Brutschrank kultiviert. Für die verschiedenen Rezipienten wurden verschiedene Temperaturen gewählt; für *E. coli* 37°C und für *Y. enterocolitica* 30°C. Am folgenden Tag wurden Verdünnungsreihen von 10⁻¹ bis 10⁻³ in PBS erstellt und vollständig ausplattiert (jeweils 450 µl). Zur Selektion der Transkonjuganten wurden verschiedene Antibiotika und Agarplatten verwendet. Die *E. coli*-Transkonjuganten wurden auf LB-Agarplatten mit Chloramphenicol und Streptomycin identifiziert, die *Y. enterocolitica* Transkonjuganten auf *Yersinia*-Selektivagar mit Chloramphenicol und *Yersinia Selective Supplement*. Die Platten wurden üN inkubiert und am nächsten Tag auf Bakterienwachstum überprüft.

6.2. Auswertung der Transkonjuganten

Wie bereits erwähnt, war der Transfer der ICE-Typ-HPI nur ausgehend von dem Region X-positiven Donor *K. pneumoniae* HU86 Cm möglich. Da die spontane Resistenzbildung bei *K. pneumoniae* sehr hoch ist, waren des Öfteren falsch-positive Klone zu finden, also HU86 Cm anstatt der gewünschten Transkonjuganten (Smith, 1976; George et al., 1995). Deswegen war es notwendig, die gewachsenen Bakterien mit verschiedenen Methoden zu identifizieren.

Die Gattung und Art der Klone wurde mittels MALDI-TOF-Analyse bestimmt. Das Vorhandensein der Chloramphenicol-resistenzmarkierten HPI wurde durch eine PCR mit den Primern HPI-fyuA-2132.for, HPI-fyuA-2947.rev und p34S-Cm2.484.rev nachgewiesen. Zuletzt wurden mittels automatisierter Analyse (BD Phoenix) die Antibiotika-Resistenzen von Donor, Rezipient und Transkonjugant ermittelt. Die erhaltenen Transkonjuganten, MG1655 ICE-Cm und MRS40 ICE-Cm, entsprechen der Gattung und Art der eingesetzten Rezipienten, haben deren Antibiotika-Resistenzen - ausgenommen Chloramphenicol - und sind HPI-positiv.

Die Berechnung der Konjugationseffizienz war nicht möglich, da sowohl durch die Konjugation mit dem Rezipienten MG1655 als auch mit MRS40 insgesamt jeweils nur ein

Transkonjugant erhalten werden konnte. Die meisten Transferversuche blieben ohne Ergebnis.

Da die ICE-Typ-HPI über die Gattungsgrenze hinaus transferiert wurde, sollte nun getestet werden, ob die PAI noch funktionsfähig ist, ob also das Siderophor Yersiniabaktin noch synthetisiert wird. Die Siderophor-Produktion kann durch den sogenannten CAS-Agar nachgewiesen werden, der sich in Anwesenheit dieser Moleküle verfärbt (Martin et al., 2013). In Abbildung 47 sind CAS-Agarplatten zu sehen, auf denen jeweils der Donor, die Rezipienten und die Transkonjuganten aufgetragen und bei 37°C und 30°C üN bebrütet wurden. Auf der Agarplatte, die bei 37°C inkubiert wurde, ist sichtbar, dass der Donor das Siderophor synthetisiert. Dies ist ebenso der Fall bei den *E. coli*-Stämmen MG1655 und MG1655 ICE-Cm, allerdings nicht bei den *Y. enterocolitica*-Stämmen MRS40 und MRS40 ICE-Cm. Der *E. coli*-Stamm MG1655 produziert das Siderophor Enterobaktin, welches wahrscheinlich für die Verfärbung verantwortlich ist (Martin et al., 2013). Bei Betrachtung der CAS-Agarplatte, die bei 30°C bebrütet wurde, synthetisieren die Stämme HU86 Cm, MG1655 und MG1655 ICE-Cm erneut Siderophore. Der Rezipient MRS40 ist wieder negativ bezüglich der Produktion, nicht aber der Transkonjugant MRS40 ICE-Cm. Dieser synthetisiert Yersiniabaktin nur bei seinem Temperaturoptimum.

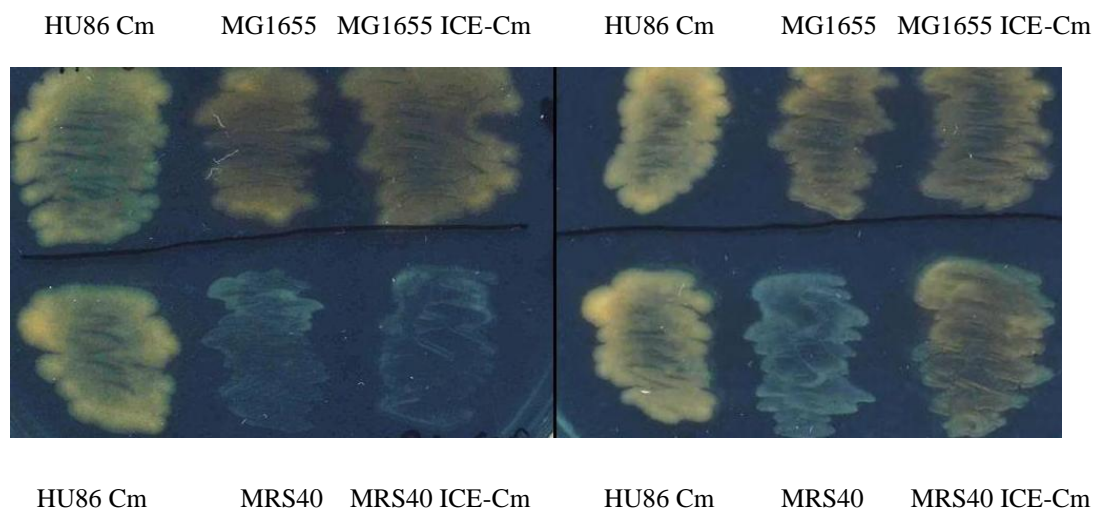


Abbildung 47: Gezeigt werden zwei CAS-Agarplatten, auf denen der Donor (HU86 Cm), die Rezipienten (MG1655 und MRS40) und die Transkonjuganten (MG1655 ICE-Cm und MRS40 ICE-Cm) üN gewachsen waren. Die Inkubationstemperaturen waren auf der linken Agarplatte 37°C und auf der rechten 30°C.

Es wurden die Integrationsorte der HPI ermittelt. Da der Transkonjugant MG1655 ICE-Cm zu einem späteren Zeitpunkt erhalten wurde, wurde dieser nicht mehr in die Sequenzierungen mit eingeschlossen. Somit sollten die *asn* tRNAs von diesem Stamm mittels PCR auf die Integration der PAI überprüft werden. Es wurde ein Primer innerhalb des ICE (insert.ICE.rev) gewählt, und vier für die tRNAs *asnT* (Ec.HPI-border.for.1998), *asnV* (attL-EC31-2633-for), *asnU* (asnW-U.25544.for) und *asnW* (Ec.rev.16500). Bei Vorhandensein der HPI sollte ein PCR-Produkt entstehen, andernfalls nicht. In Abbildung 48 ist ein vereinfachtes Schema der PCR zu sehen.

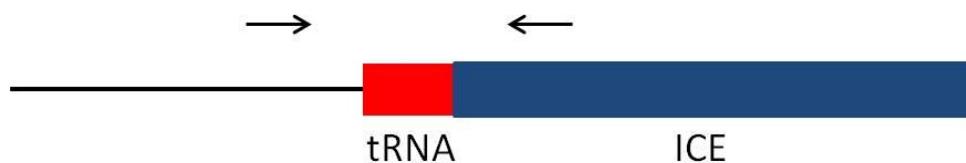


Abbildung 48: Vereinfachtes Schema der PCR zur Untersuchung der Integration der ICE-Typ-HPI in die *asn* tRNAs im Transkonjuganten *E. coli* MG1655 ICE-Cm. Der schwarze Strich stellt das Genom *upstream* der jeweiligen *asn* tRNA dar. Die tRNA ist in Rot gekennzeichnet, das ICE in Blau. Die verwendeten Primer sind als Pfeile dargestellt. Es wurde ein Primer innerhalb des ICE gewählt, und jeweils ein Primer *upstream* der tRNA.

Es wurden von diesem Transkonjuganten jeweils sechs Einzelkolonien für jede untersuchte tRNA verwendet. Als Negativkontrolle wurde der Rezipient MG1655 verwendet. Das Gelbild ist in Abbildung 49 zu sehen. Es ist zu erkennen, dass die ICE-Typ-HPI dreimal in das Genom des *E. coli*-Stammes integriert wurde, und zwar in die tRNAs *asnT*, *asnU* und *asnW*, nicht aber in *asnV*. Mit der Negativkontrolle wurde kein PCR-Produkt erzeugt.

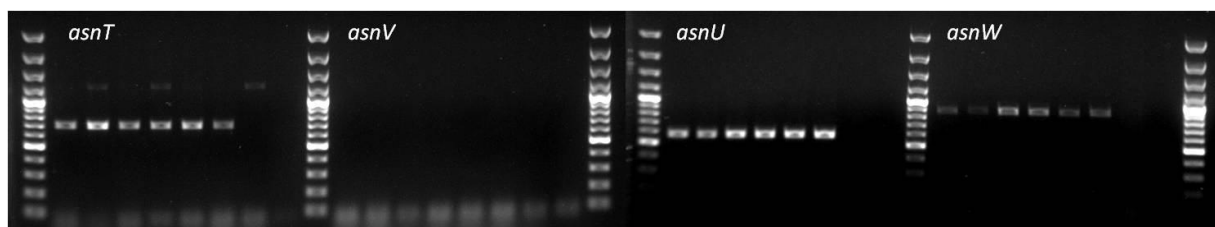


Abbildung 49: Gelbild der PCR zur Untersuchung der Integration der ICE-Typ-HPI in die *asn* tRNAs des Transkonjuganten MG1655 ICE-Cm. Es wurden für jede tRNA sechs Einzelkolonien des Stammes verwendet. Die Integration erfolgte in die tRNAs *asnT*, *asnU* und *asnW*. Als Negativkontrolle wurde der Rezipient MG1655 verwendet. Die letzte Position jeder einzelnen PCR war eine Leerkontrolle. Als Marker wurde der GeneRuler™ 100 bp verwendet.

Da der Transkonjugant MRS40 ICE-Cm zuerst erhalten wurde, wurde dieser noch komplett sequenziert. Die bioinformatische Auswertung ergab, dass nur eine ICE-Typ-HPI übertragen wurde, und zwar komplett und ohne zusätzliche Donoren-DNA. Der Integrationsort der PAI wurde ebenfalls *in silico* bestimmt und mit dem Referenzstamm *Y. enterocolitica* 8081 (NCBI-Zugangsnummer NC_008800), der eine immobile HPI besitzt, mit einem *Alignment* abgeglichen. Es wurden die drei Stämme MRS40 ICE-Cm, 8081 und HU86 Cm miteinander verglichen. Die Integration fand in der *asn* tRNA YEt054 entsprechend dem Stamm 8081 statt, in der auch dessen HPI integriert ist (Thomson et al., 2006).

D. Diskussion

Extraintestinal pathogene *Escherichia coli* (ExPEC) sind für viele Infektionen außerhalb des Magen-Darm-Trakts verantwortlich. Diese können recht harmlos, aber auch lebensbedrohlich ausfallen (Johnson und Russo, 2002). Die Eigenschaften, die ExPECs zu erfolgreichen Infektionen befähigen, sind oft auf genomischen Inseln (GEIs) codiert. Dazu gehören z.B. Antibiotikaresistenzen, aber auch Systeme zur Eisenaufnahme und zur Unterdrückung der Immunantwort (Dobrindt et al., 2004). GEIs können über horizontalen Gentransfer (HGT) an andere Bakterien weitergegeben werden, wodurch deren Virulenz steigen kann. Durch diese Entwicklung sind ExPECs aus *E. coli*-Stämmen entstanden (Dobrindt und Hacker, 2001). Die Weitergabe von GEIs durch HGT geschieht meist durch Konjugation; ein Prozess, bei dem ein Donor einen Teil seiner genetischen Informationen auf einen Rezipienten überträgt. Pathogenitätsinseln (PAIs) sind eine Untergruppe der GEIs, die Virulenzgene tragen. Es wurde bereits eine Vielzahl an PAIs beschrieben (Dobrindt et al., 2004; Schmidt und Hensel, 2004). Durch das Vorhandensein von Mobilitäts- und Transfer-Genen sind manche PAIs noch aktiv übertragbar. Allerdings existieren in der Art *E. coli* vor allem Inseln, die diese Gene nicht besitzen oder durch eine Deletion verloren haben. Diese sind dann aktiv nicht mehr transferierbar (Schubert et al., 2004a).

Die Fragestellung dieser Arbeit befasst sich mit den Übertragungsmechanismen von sowohl aktiv als auch passiv transferierbaren PAIs. Im Fokus steht die *High-Pathogenicity Island* (HPI), die für die Synthese des Siderophors Yersiniabaktin codiert. Von dieser Insel gibt es zwei Varianten: eine aktiv übertragbare Form, die ICE-Typ-HPI, und eine kurze Form, die durch eine Deletion die Transfer- und Mobilitäts-Gene verloren hat (Schubert et al., 2004a). Zudem werden noch drei weitere PAIs untersucht, wovon zwei Inseln direkt neben der kurzen HPI im *E. coli*-Genom liegen. Dies sind die *pks*- und die *serU*-Insel, deren Produkte mit eukaryotischen Wirtszellen interagieren können (Putze et al., 2009; Schubert et al., 2010). Die letzte Insel, die PAI III, ist weit von den bereits erwähnten Inseln entfernt und codiert für die Synthese des Siderophors Salmochelin (Russo et al., 1999). Die Übertragungsmechanismen werden auf zwei verschiedenen Wegen untersucht. Einerseits werden phylogenetische Analysen der PAIs *in silico* mit einer großen Stammsammlung durchgeführt, andererseits die aktiven und passiven Transfers der Inseln *in vitro* nachgestellt.

1. Next-Generation Sequencing

Das *Next-Generation Sequencing* (NGS) hat die Forschung und die Medizin revolutioniert. Während Sequenzier-Einheiten der ersten Generation sehr zeit- und kostenaufwendig waren, generieren die der nächsten Generation eine so enorme Menge an Rohdaten, dass die neuen Herausforderungen darin liegen, genug Speicherplatz für die Daten und Personal für die Auswertung der Sequenzen bereit stellen zu können. Die Entwicklung der Sequenzierer wird auch weiterhin in Richtung Kostensenkung und Erhöhung des Durchsatzes gehen. Somit können nun auch große Stammsammlungen mit einem machbaren Aufwand entschlüsselt und für die Wissenschaft zugänglich gemacht werden (Frelinger, 2015).

Die *Escherichia coli* Referenz (ECOR)-Sammlung wurde in den 1980er gesammelt und seitdem von vielen forschenden Arbeitsgruppen verwendet. Allerdings wurde die ECOR-Sammlung nie vollständig sequenziert, obwohl bereits Unterschiede bei manchen Isolaten beim Vergleich der Sammlungen berichtet wurden (Johnson et al., 2001a). Durch die Verwendung von NGS wird es möglich, dass einzelne Forschungsgruppen ihre eigenen Sammlungen entschlüsseln und somit auch besser analysieren können.

In dieser Arbeit wurden insgesamt 100 bakterielle Genome mit der NGS-Technologie von Illumina sequenziert, um die Mechanismen des HGT zu untersuchen. Für phylogenetische Studien innerhalb der Art *E. coli* wurde zum ersten Mal die komplette ECOR-Sammlung sequenziert. Die Qualität der Rohdaten ist recht gut, was auch an dem hohen Phred-Qualitätswert von 40 zu sehen ist. Dieser Wert beschreibt, dass die Wahrscheinlichkeit einer falsch eingebauten Base bei der Sequenzierung im Durchschnitt bei 0,001% liegt (Ewing und Green, 1998; Ewing et al., 1998). Die Aussage, dass 99,84% der Rohdaten keine mehrdeutigen Basen enthalten, bekräftigt zudem die Qualität. Es konnte auch die Entstehung von Hybrid-Genomen durch die Aufnahme großer DNA-Abschnitte mittels NGS verfolgt werden. Dies war durch die teilweise sehr langen *de novo* assemblierten *Contigs* von bis zu 400 kb möglich.

Allerdings konnten nicht alle Analysen aufgrund fehlender Sequenzabschnitte durchgeführt werden. Unter anderem sind nicht alle Gene für die Erstellung einer Multilocus-Sequenz-Typisierung (MLST) nach dem Achtman-Schema vorhanden. Zudem fehlen große Abschnitte in dem Bereich der PAI III. Dieses Problem könnte mitunter durch die Verwendung von längeren *Read*-Längen behoben werden, um größere *Contigs* und damit weniger Lücken zu erhalten. Auch ein höherer Durchsatz, also die Generierung von mehr *Reads*, würde zu mehr

genetischen Informationen führen. Trotzdem ist es möglich, mit den unvollständigen Genomen tiefere Einblicke in den HGT von PAIs zu erhalten.

2. Vergleich der PCR Ergebnisse

Um im Vorfeld einen Überblick über die untersuchten Bakterienstämme zu bekommen, wurden diese mit verschiedenen Polymerase-Kettenreaktionen (PCR) charakterisiert (Tab. 13). Der Fokus liegt hier vor allem auf den vorhandenen PAIs und auf den phylogenetischen Gruppen der *E. coli*-Stämme.

Die Verteilung der vier PAIs HPI, *pks*-Insel, *serU*-Insel und PAI III stimmt größtenteils mit bereits publizierten Daten überein (Martin et al., 2013; Schubert et al., 2010). Es wurden Unterschiede in vier Stämmen der ECOR-Sammlung festgestellt. In der Sammlung, die für diese Arbeit verwendet wird, besitzt der Stamm ECOR43 eine HPI anstatt HPI-negativ zu sein, der Stamm ECOR57 hat keine PAI III im Gegensatz zur veröffentlichten Variante. Während publiziert wurde, dass der Stamm ECOR64 im Besitz der HPI, der *pks*-Insel und der PAI III ist, wurde in dem hier verwendeten Isolat nur die HPI gefunden. Der letzte Unterschied liegt darin, dass der Stamm ECOR70 in dieser Sammlung eine PAI III besitzt, im Gegensatz zu einer Publikation (Martin et al., 2013). Der Fakt, dass der Stamm ECOR70 nicht in allen ECOR-Sammlungen eine PAI III besitzt, könnte daran liegen, dass der durch eine PCR nachgewiesene *iroBCDEN*-Gen-Cluster bei diesem Isolat auf einem Plasmid lokalisiert ist. Da Stämme leicht Plasmide aufnehmen oder verlieren können, wäre dies eine mögliche Erklärung. Aufgrund der Aussage, dass die PAI III sonst nur in der phylogenetischen Gruppe B2 zu finden ist, bekräftigt dies die Annahme, dass dieser B1-Stamm ein Plasmid aufgenommen haben könnte (Dobrindt et al., 2002). Ansonsten zeigen die Unterschiede erneut, dass die ECOR-Sammlungen von unterschiedlichen Arbeitsgruppen variieren können (Johnson et al., 2001a).

Die Einteilung der Art *E. coli* in die vier verschiedenen phylogenetischen Hauptgruppen A, B1, B2 und D ist mit verschiedenen Methoden möglich, mitunter durch eine Triplex-PCR (Clermont et al., 2000). Diese wurde bei allen *E. coli*-Isolaten vor den eigentlichen Untersuchungen durchgeführt, um die Ergebnisse mit den *in silico*-Daten vergleichen zu können. Im Vergleich mit bereits publizierten Daten wurde nur ein Unterschied festgestellt (Herzer et al., 1990; Martin et al., 2013). Der Stamm ECOR43, der in dieser Sammlung auch

eine HPI besitzt, ist in der phylogenetischen Hauptgruppe A zu finden, was auch Clermont *et al.* durch die Verwendung der Triplex-PCR feststellten. Es wurde beschrieben, dass dieser Stamm entweder nicht zuzuordnen ist oder in der Untergruppe E beheimatet ist. Drei weitere Stämme, die anhand der Triplex-PCR in die Hauptgruppe D eingeordnet werden, waren ebenfalls in der erwähnten Publikation in keine Gruppe einzuordnen bzw. in der Untergruppe E. Dabei handelt es sich um die Stämme ECOR31, ECOR37 und ECOR42 (Clermont et al., 2000; Martin et al., 2013). Diese drei Isolate bilden auch in den phylogenetischen Stammbäumen der *Housekeeping*-Gene eine eigene Gruppe (Abb. 17 - 19). Die Triplex-PCR eignet sich somit nur zur raschen Einteilung in die vier Hauptgruppen und kann nicht den Verwandtschaftsgrad der untersuchten Stämme wiedergeben.

3. Housekeeping-Gene und phylogenetische Gruppen

Die Einteilung in die phylogenetischen Gruppen wird vor allem durch die MLST von Fragmenten verschiedener *Housekeeping*-Gene vollzogen. Dafür werden in der Regel mehrere PCRs pro Isolat durchgeführt, deren Produkte anschließend sequenziert werden. Die einzelnen Sequenzen eines Stammes werden danach *in silico* aneinandergereiht und anhand eines erstellten Stammbaumes mit anderen bakteriellen Sequenzen verglichen (Wirth et al., 2006; Jaureguy et al., 2008).

Mit der NGS-Technik ist es möglich, den recht großen Zeitaufwand für die Durchführung der PCRs zu umgehen, da diese *in silico* durchgeführt werden können (Larsen et al., 2012). Zusätzlich können neben den Fragmenten auch die gesamten *Housekeeping*-Gene analysiert werden. In dieser Arbeit wurde das Pasteur-Schema für die MLST angewandt (Jaureguy et al., 2008). Dieses vergleicht die Fragmente von acht Genen (*dinB*, *trpA*, *trpB*, *pabB*, *putP*, *icdA*, *polB* und *uidA*). Es wurde zusätzlich auch eine modifizierte Variante unter der Verwendung von sechs *Housekeeping*-Gen-Fragmenten (*trpA*, *trpB*, *pabB*, *putP*, *icdA* und *polB*) angewandt (Escobar-Paramo et al., 2004).

In dieser Studie wurden sowohl die gesamten sechs Gene als auch deren Fragmente analysiert. Zusätzlich wurden die Fragmente der acht *Housekeeping*-Gene verwendet, allerdings war eine Untersuchung der gesamten acht Gene aufgrund fehlender Sequenzabschnitte nicht möglich. In den drei erstellten Stammbäumen (Abb. 17 - 19) werden

die Bakterienstämme größtenteils in die phylogenetischen Gruppen eingeteilt, die auch durch die Triplex-PCR bestimmt wurden.

In allen Bäumen trennt sich die phylogenetische Gruppe B2 deutlich von den anderen Gruppen ab und ist am eigenständigsten. Die Stämme dieser Gruppe sind alle zumindest im Besitz der HPI. Dies zeigt erneut die erhöhte Virulenz der B2-Isolate innerhalb der Pathogruppe ExPEC (Clermont et al., 2000).

Die phylogenetische Hauptgruppe D ist im Gegensatz zur Gruppe B2 sehr divergent, was auch anhand der Homologie-Werte sichtbar ist (Tab. 16). Es wurde auch beschrieben, dass die Mutationsrate in der phylogenetischen Hauptgruppe D am höchsten ist (Wirth et al., 2006). Es bilden sich in allen Stammbäumen drei sichtbare Untergruppen innerhalb der Hauptgruppe D, und in der Analyse der Fragmente (Abb. 17 und 18) ist zusätzlich der B1-Stamm ECOR30 innerhalb dieser Gruppe zu finden. Es wurde bereits beschrieben, dass neben den vier phylogenetischen Hauptgruppen auch Untergruppen zu finden sind (Jaureguy et al., 2008). Zu der Untergruppe E gehören Stämme, denen in früheren Studien keine phylogenetische Gruppe zugeteilt werden konnte (Herzer et al., 1990). In diese Untergruppe wurden die Stämme ECOR31, ECOR37, ECOR42 und ECOR43 eingeteilt (Martin et al., 2013). Diese Isolate bilden mit Ausnahme des Stammes ECOR43, der in die phylogenetische Gruppe A eingeteilt wird, auch in dieser Arbeit eine Untergruppe innerhalb der Hauptgruppe D, die hier als E bezeichnet werden kann. Die beschriebene Untergruppe F beinhaltet Stämme, die aufgrund der Triplex-PCR in die Gruppe D eingeteilt werden, sich aber im Stammbaum von deren Mitgliedern distanzieren (Jaureguy et al., 2008). Dies erklärt die beiden verbleibenden Gruppen, die sich demnach in die Hauptgruppe D und die Nebengruppe F aufteilen.

Der B1-Stamm ECOR30 ist in den Fragment-basierten Stammbäumen (Abb. 17 und 18) innerhalb der Gruppe D und in der Analyse der sechs vollständigen *Housekeeping*-Gene (Abb. 19) in der Gruppe B1 zu finden. Unter der Verwendung des Achtman-MLST-Schemas wurde eine sogenannte ABD-Gruppe beschrieben, die durch vermehrte Rekombination nun DNA-Abschnitte von allen phylogenetischen Gruppen enthält (Wirth et al., 2006). Aufgrund der unterschiedlichen Einteilungen in die verschiedenen Gruppen könnte der Stamm ECOR30 in diese Gruppe fallen. Dies würde auch die entferntere Verwandtschaft zu den übrigen B1-Stämmen in der Analyse der sechs vollständigen *Housekeeping*-Gene erklären (Abb. 19).

Die phylogenetischen Gruppen A und B1 sind in allen Stammbäumen sehr verwandt, trennen sich aber dennoch größtenteils in ihre eigenen Gruppen auf. Dies liegt daran, dass diese

beiden Hauptgruppen sogenannte Schwester-Gruppen sind, die sich erst im Lauf der Evolution voneinander abgespalten (Clermont et al., 2000; Wirth et al., 2006). Die Tatsache, dass sich die beiden B1-Stämme ECOR70 und ECOR72 in den Fragment-basierten Stammbäumen (Abb. 17 und 18) zwischen den Gruppen B1 und A anlagern, kann mit der früheren Verwandtschaft der beiden phylogenetischen Gruppen erklärt werden. Es wurde auch berichtet, dass die Gruppe A paraphyletisch ist, da sich manche B1-Stämme näher zu dieser Gruppe einordnen als zu der eigenen (Jaureguy et al., 2008).

Durch die Analyse der sechs vollständigen *Housekeeping*-Gene (Abb. 19) wird eine AxB1 Mischgruppe sichtbar. Diese vier Stämme, von denen durch die Triplex-PCR zwei in die Gruppe A (ECOR16 und ECOR24) und zwei in die Gruppe B1 (ECOR69 und ECOR70) eingeordnet werden, haben DNA-Abschnitte von der jeweils anderen phylogenetischen Gruppe. Eine AxB1-Gruppe wurde auch unter Verwendung des Achtman-MLST-Schemas beobachtet (Wirth et al., 2006). Bei beiden A-Stämmen sind die rekombinanten DNA-Abschnitte in den Genen *pabB* und *putP* zu finden, zudem sind die Abschnitte auch an denselben Stellen lokalisiert. Die betroffenen *Housekeeping*-Gene der B1-Stämme sind *putP* und *icdA*, wobei nur im Letzteren die rekombinanten Abschnitte innerhalb der beiden Isolate an derselben Stelle lokalisiert sind. Der Stamm ECOR70 hat eine zusätzliche Durchmischung mit DNA der phylogenetischen Gruppe A in dem Gen *polB* (Abb. 20). Während die beiden A-Stämme wahrscheinlich derselben Rekombination unterlagen, waren bei den beiden B1-Stämmen anscheinend verschiedene Rekombinationsereignisse eingetreten. Bezüglich des Stammes ECOR70 wurde bereits vermutet, dass dieser ein Hybrid-Stamm sein könnte, da sowohl DNA der phylogenetischen Gruppe A als auch der Gruppe B1 in einigen *Housekeeping*-Genen gefunden wurde (Clermont et al., 2000). Das Gen *icdA*, das von beiden MLST-Schemata verwendet wird, hatte in der Studie, die das Achtman-Schema verwendet, eine erhöhte Rekombinationsrate in den Gruppen A, B1 und D (Wirth et al., 2006). Im Vergleich dazu hat das Gen *icdA* in dieser Arbeit nur eine verhältnismäßig geringe Durchmischung mit DNA der anderen phylogenetischen Gruppen, und auch nur in den getesteten B1-Stämmen ECOR69 und ECOR70.

Zur Untersuchung des HGT von PAIs war es zudem wichtig, die analysierten Stämme auf Klonalität zu prüfen. Um eine Unterscheidung zwischen vertikalem und horizontalem Transfer machen zu können, wurde untersucht, ob die Mitglieder der PAI-Gruppen voneinander abstammen und die Inseln über Zellteilung weitergegeben wurden, oder ob die Übertragung durch HGT stattgefunden hat. Das gewählte Pasteur-Schema ist hierfür

besonders geeignet, da die verwendeten *Housekeeping*-Gene unter geringem Einfluss von HGT stehen sollen, und somit die stammesgeschichtliche Evolution darstellen (Lecointre et al., 1998; Denamur et al., 2000). Es wurde ein Stammbaum mit den Sequenzen der sechs vollständigen *Housekeeping*-Gene aller HPI-positiven Stämme erstellt (Abb. 21). Die Anordnung der Isolate wird vor allem durch die phylogenetischen Gruppen bestimmt. Innerhalb der Gruppe B2 sind die Stämme der PAI-Gruppe 3 gleichmäßig verteilt. Die beiden Stämme der PAI-Gruppe 2b, S107 und S108, haben anhand des Stammbaums eine höhere Verwandtschaft. Dies wurde auch in einem bereits publizierten Stammbaum gezeigt, allerdings wurden durch weitere Untersuchungen Unterschiede in dem Vorhandensein von gewissen genomischen Regionen zwischen diesen beiden Stämmen festgestellt (Le et al., 2007). Innerhalb der PAI-Gruppe 2a sind die Stämme, ausgenommen ECOR65, näher miteinander verwandt. Es wurden allerdings Unterschiede in der Verteilung der PAI III gefunden, da nicht alle Isolate dieser PAI-Gruppe eine PAI III besitzen. Zudem gilt die ECOR-Sammlung als repräsentativ für die Diversität der Art *E. coli*, da sie von verschiedenen Wirten und Kontinenten isoliert wurde (Ochman und Selander, 1984). Somit sollte diese nähere Verwandtschaft nicht aufgrund von Klonalität entstanden sein. Der Stamm ECOR65 ist näher verwandt mit der PAI-Gruppe 3 als mit seiner eigentlichen, PAI-Gruppe 2a. Das Fazit dieser Analyse ist, dass die Übertragung der PAIs durch HGT stattgefunden hat.

Abschließend kann gesagt werden, dass durch die Nutzung der NGS-Daten eine schnellere Erstellung von MLST-basierten Stammbäumen möglich ist. Durch den Erhalt von mehr genetischen Informationen durch die vollständigen Gensequenzen des Pasteur-MLST-Schemas sind zusätzlich tiefere Einblicke in die Phylogenie der Art *E. coli* möglich. Durch die Analyse der vollständigen Gene können Ergebnisse erzielt werden, die auch anhand des Achtman-MLST-Schemas gezeigt wurden. In den meisten Fällen ordnet der Stammbaum der vollständigen Gene die Stämme in die phylogenetischen Gruppen ein, in denen diese auch in den Fragment-basierten Stammbäumen eingeordnet werden, was auch zu erwarten ist. Zudem können keine relevanten Unterschiede zwischen der Verwendung von sechs oder acht Gen-Fragmenten festgestellt werden.

4. Pathogenitätsinseln

Viele Virulenzfaktoren wie Toxine und Eisenaufnahmesysteme sind auf PAIs lokalisiert. Diese Inseln wurden zuerst in humanpathogenen *E. coli* beschrieben, wurden aber mittlerweile auch in anderen pathogenen Arten entdeckt (Hacker und Kaper, 2000). In dieser Arbeit wurde die Phylogenie von immobilisierten PAIs der Art *E. coli* analysiert, und zwar die der HPI, der *pks*-Insel, der *serU*-Insel und der PAI III (Schubert et al., 2009; Homburg et al., 2007; Cirl et al., 2008; Dobrindt et al., 2002). Da von der HPI auch ein aktiv transferierbarer ICE-Typ in einigen Isolaten der Familie *Enterobacteriaceae* existiert, wurde diese als Beispiel für mobile PAIs in dieser Familie gewählt (Schubert et al., 2004a; Putze et al., 2009).

4.1. Phylogenie und Transfer der immobilisierten Pathogenitätsinseln

Da der Transfer von immobilisierten PAIs nicht mehr aktiv stattfinden kann, muss die Verbreitung entweder passiv abgelaufen sein oder kann gar nicht mehr vollzogen werden. Dass die Übertragung nur noch vertikal verläuft, kann durch die Analyse der *Housekeeping*-Gene größtenteils ausgeschlossen werden. Es bestand noch die Möglichkeit, dass beim Eintreffen der Inseln in die Art *E. coli* viele Transfer-Events stattgefunden haben und sich die Inseln in bestimmten Stämmen etabliert haben, die heutzutage noch zu finden sind. Um dies zu untersuchen, wurden von allen Inseln Stammbäume erstellt und miteinander verglichen. Zudem wurden von den drei PAIs HPI, *pks*- und *serU*-Insel jeweils sechs codierende Gene untersucht, um Rückschlüsse auf den Grad der Konservierung und daraus folgend auf das Alter machen zu können. Es wurde bereits berichtet, dass sowohl die *pks*-Insel als auch die *serU*-Insel in der Art *E. coli* nur in der phylogenetischen Gruppe B2 und nur in Verbindung mit der HPI gefunden wurde, mit Ausnahme von zwei B1-Stämmen mit einer *pks*-Insel innerhalb der Region III einer ICE-Typ-HPI (Putze et al., 2009; Schubert et al., 2010). Um die Übertragung von mehreren Inseln *en bloc* zu analysieren, wurden die einzelnen PAI-Gruppen bezüglich ihrer Insel-Subtypen betrachtet. Zusätzlich wurde das *Backbone*-Genom zwischen den PAIs, die inter-PAI-Regionen, untersucht.

Der Transfer der immobilisierten PAIs HPI, *pks*- und *serU*-Insel wurde zur Bestätigung der *in silico*-Ergebnisse, die einen *en bloc*-Transfer von bestimmten Inseln (PAI-Gruppen) zeigen, durchgeführt. Hierfür wurden zwei *E. coli*-Donoren erstellt, die entweder im Besitz der HPI und der *pks*-Insel (NU14 Cm F') oder von allen drei genannten PAIs (536 Cm F') sind (Johnson et al., 2001b; Hacker et al., 1992). Die Konjugation der PAIs erfolgte mithilfe eines

F'-Plasmides, welches neben den Inseln noch das umliegende *Backbone* des Donors überträgt (Griffiths et al., 1999). Das Ausmaß des Transfers wurde durch NGS sichtbar gemacht.

Ein interessantes Ergebnis ist, dass die Stämme, die eine *serU*- und/oder *pks*-Insel besitzen, nur aus den Wirten Mensch und Affe und aus den Ländern Schweden, USA und Deutschland isoliert wurden (Le et al., 2007). Die Bakterien mit diesen PAIs scheinen auf bestimmte Wirte und deren Immunsystem angepasst zu sein und häufig in Industrieländern vorzukommen.

4.1.1. Die Untersuchung einer gesamten PAI im Gegensatz zu einem MLST-Ansatz bringt neue Informationen bei "jüngeren" Inseln

Bei dem Vergleich der Stammbäume der gesamten PAIs mit den zugehörigen sechs codierenden Genen können unterschiedliche Beobachtungen gemacht werden. Während in den beiden HPI-Stammbäumen (Abb. 22 und 23) kaum Veränderungen festzustellen sind, kann bei der *serU*-Insel unter Verwendung von mehr genetischen Informationen eine deutlichere Auftrennung innerhalb der PAI-Gruppe 3 erzielt werden (Abb. 26 und 27). Aufgrund der absolut identischen sechs Gene konnte für die *pks*-Insel kein MLST-basierender Stammbaum erstellt werden. Dies spiegelt auch das Alter der PAIs wieder. Während die HPI wohl als erste Insel in die Art *E. coli* eintraf und auf alle phylogenetischen Gruppen verteilt wurde, kamen die *serU*- und die *pks*-Insel erst später hinzu. Da die codierenden Gene dieser beiden Inseln noch hoch konserviert sind, scheinen die auftretenden SNPs vor allem in den nicht-codierenden Bereichen zwischen den Genen zu liegen, was für jüngere PAIs sprechen würde (Putze et al., 2009; Schubert et al., 2010). Bei der HPI scheinen die SNPs schon in den codierenden Genen fest "verankert" zu sein, was für eine ältere Insel spricht. Bei Betrachtung des Maßstabes, der in dem Stammbaum der *pks*-Insel (Abb. 25) die Anzahl an SNPs pro Base angibt, scheint diese PAI zuletzt in die Art *E. coli* übertragen worden zu sein. Zusätzlich ist zu erwähnen, dass die untersuchten sechs Gene des MLST-Stammbaums der HPI aufgrund ihrer Größe einen erheblichen Anteil der eigentlichen Insel ausmachen und deswegen wohl auch die Phylogenie der gesamten Insel widerspiegeln.

4.1.2. Die PAIs HPI, *pks*- und *serU*-Insel kamen unabhängig in die Art *E. coli* und wurden anschließend *en bloc* innerhalb der phylogenetischen Gruppe B2 weitergegeben

Im Gegensatz zu der *pks*- und der *serU*-Insel, die nur in Bakterienstämmen der phylogenetischen Gruppe B2 zu finden sind, ist die HPI in allen Gruppen präsent. Die Sequenzen der HPIs ordnen sich größtenteils entsprechend der dazugehörigen phylogenetischen Gruppen an, mit Durchmischungen sowohl der Gruppen A und B1 als auch B2 und D (Abb. 22). Diese HPI-Sequenzen könnten sich in mehr und weniger pathogene Insel-Subtypen innerhalb der ExPEC aufteilen, da auch manche Kommensale bekanntermaßen eine HPI besitzen, diese aber eher die Fitness anstatt der Virulenz erhöht (Dobrindt et al., 2004; Clermont et al., 2000).

Anhand des HPI-Stammbaumes (Abb. 22) kann gezeigt werden, dass die *pks*- und die *serU*-Insel nicht *en bloc* in die Art *E. coli* gelangten, da diese beiden Inseln sonst gemeinsam auftreten müssten. Auch ein anschließender Verlust kann hier wahrscheinlich ausgeschlossen werden, da sich die PAI-Gruppen 2a und 2b sonst einen gemeinsamen oder nahe verwandten HPI-Subtypen teilen würden. Es bilden sich in den drei Stammbäumen der HPI, *pks*- und *serU*-Insel (Abb. 22, 25 und 26) für die jeweiligen PAI-Gruppen spezifische Insel-Subtypen. Zudem sind die inter-PAI-Regionen innerhalb der PAI-Gruppen nahezu identisch. Zusammengefasst bedeutet dies, dass nach dem unabhängigen Eintreffen der Inseln HPI, *pks*- und *serU*-Insel diese gemeinsam mit dem dazwischenliegenden *Backbone en bloc* über HGT auf weitere *E. coli*-Stämme übertragen wurden. Die *in vitro*-Nachstellung des Transfers bestätigt eine *en bloc*-Übertragung.

Die *in silico*-Untersuchung der sequenzierten Transkonjuganten zeigt, dass die PAIs vollständig übertragen werden. Es müssen homologe Bereiche innerhalb der auf dem F'-Plasmid vorliegenden Donor-DNA und der genomischen Rezipienten-DNA für eine homologe Rekombination via *Crossing-Over* vorhanden sein (Adelberg und Pittard, 1965). Da der Rezipient selbst keine PAIs besitzt, konnte in diesen auch keine Rekombination stattfinden. In jeden Transkonjuganten wurde von dem F'-Plasmid jeweils nur ein zusammenhängender Teil an DNA integriert und nicht mehrere an verschiedenen Stellen. Dafür hätten mindestens vier anstatt zwei *Crossing-Over* stattfinden müssen, was sehr selten geschieht (Griffiths et al., 1999).

4.1.3. Die PAI III wurde nicht *en bloc* mit den PAIs HPI, *pks*- und *serU*-Insel übertragen

Die PAI III konnte nicht vollständig analysiert werden, da große Sequenz-Abschnitte dieser Insel in einigen Stämmen fehlen. Dies könnte durch die Überreste von IS-Elementen und Transposons zustande gekommen sein (Dobrindt et al., 2002). Es wurde ein 10,5 kb großer Bereich inklusive des *iro*-Gen-Clusters gefunden, der in allen PAI III-positiven Isolaten vorhanden ist. Die Sequenzen dieses Abschnitts wurden anhand eines Stammbaumes verglichen (Abb. 31). Mehr als die Hälfte der Isolate sind bezüglich ihrer Sequenzen fast komplett identisch, während zwei Stämme, ECOR62 und ECOR70, einen weit entfernten Cluster bilden. Durch das Hinzufügen der auf 10,5 kb getrimmten Sequenz des *E. coli*-Plasmids p300 wird ersichtlich, dass die *iroBCDEN*-Gene womöglich extrachromosomal vorliegen (Sorsa et al., 2003). Um dies zu überprüfen, wurden *in silico* die Genomregionen *up*- und *downstream* der *thrW* tRNA der Stämme ECOR62 und ECOR70 untersucht (Abb. 30). In dieser tRNA ist die genomische PAI III innerhalb der Art *E. coli* integriert (Dobrindt et al., 2002). Es kann gezeigt werden, dass in diesem Bereich keine PAI inseriert ist und diese also auf einem Plasmid vorliegen muss.

Bei der Analyse der Bereiche *up*- und *downstream* der PAI III wird zuerst ersichtlich, dass sich die Sequenzen des *Backbone*-Genoms vor allem entsprechend der phylogenetischen Gruppen anordnen, da sich die Sequenz des B1-Stamms ECOR70 deutlich von den Sequenzen der anderen untersuchten B2-Stämme unterscheidet (Abb. 32 und 34). Durch einen Vergleich innerhalb der phylogenetischen Gruppe B1 kann gezeigt werden, dass der Stamm ECOR70 auch große Unterschiede zu den anderen B1-Stämmen aufweist (Abb. 33 und 36). Es wurde bereits erwähnt, dass dieser Stamm ein Hybrid-Genom aufweist, da in *Housekeeping*-Genen von ECOR70 genetisches Material von verschiedenen phylogenetischen Gruppen gefunden wurde (Clermont et al., 2000). Dies wird nun nicht nur durch die Untersuchung der *Housekeeping*-Gene gezeigt, sondern auch durch die Auswertung der *up*- und *downstream*-Bereiche der *thrW* tRNA. Der B1-Stamm ECOR69, der sich ebenfalls nicht zu den anderen Vertretern dieser phylogenetischen Gruppe einreicht, zeigt auch eine ähnliche Durchmischung der *Housekeeping*-Gene (AxB1-Mischgruppe) wie der Stamm ECOR70 (Abb. 20). Dies deutet ebenso auf ein hybrides Genom von ECOR69 hin.

Die PAI-Gruppen 1 und 3 bilden in keinem der drei untersuchten Abschnitte - Teilstück der PAI III, *up*- und *downstream*-Bereich - einen gemeinsamen Cluster (Abb. 31, 32 und 34). Die Sequenzen der PAI-Gruppe 2a sind bezüglich des PAI III-Teilstücks und des *upstream*-

Bereichs identisch, aber nicht bezüglich des *downstream*-Bereichs. Deswegen kann davon ausgegangen werden, dass die PAI III nicht *en bloc* mit den anderen untersuchten PAIs übertragen wurde.

Auch bei der Übertragung der Inseln mit einem F'-Plasmid konnte die PAI III bei allen *in vitro* untersuchten Transkonjuganten weder per PCR noch bei den sequenzierten Isolaten *in silico* nachgewiesen werden. Da im Genom des *E. coli*-Stammes 536 diese Insel mehr als 1,5 Mb entfernt von der *serU*-Insel lokalisiert ist, erscheint ein *en bloc*-Transfer durch ein F'-Plasmid gemeinsam mit den PAIs HPI, *pks*- und *serU*-Insel als eher unwahrscheinlich.

4.1.4. Die HPIs der Stämme ECOR31 und ECOR72 spiegeln den Übergang der ICE-Typ-HPI zu der kurzen *E. coli*-HPI wieder

Die Stämme ECOR31 und ECOR72 bilden in dem HPI-Stammbaum (Abb. 22) einen eigenen Cluster, der allerdings in sich keine hohe Verwandtschaft zeigt. Bei der Untersuchung der inter-PAI-Regionen sind diese beiden Stämme unter den HPI-negativen Isolaten zu finden (Abb. 28 und 29). Der Stamm ECOR31 besitzt bekanntermaßen eine ICE-Typ-HPI, die für diese Untersuchungen auf die Länge der deletierten HPIs gekürzt wurde. Da das ICE ohne benachbartes *Backbone*-Genom in das Chromosom integriert wird, erklärt dies die Einordnung zu HPI-negativen Stämmen in den Stammbäumen der inter-PAI-Regionen (Schubert et al., 2004a). Die HPI des Stammes ECOR72 wurde als mögliche intermediäre Form zwischen der ECOR31-HPI und dem weit verbreiteten kurzen Typ beschrieben (Schubert et al., 2009). Dies erklärt die Cluster-Bildung mit dem Stamm ECOR31 in dem HPI-Stammbaum (Abb. 22). Bei dem Vergleich der kurzen HPI mit der Kern-Region des ICE-Typs ordnet sich die ECOR72-Sequenz zu den ICE-Sequenzen an (Abb. 43), was die Hypothese zusätzlich bekräftigt, dass es sich bei der ECOR72-HPI um eine Übergangsform zwischen ICE-Typ und normaler HPI handelt. Die Zuordnung der ECOR72-Sequenz zu denen der HPI-negativen Stämme in der Analyse der inter-PAI-Regionen kann damit begründet werden.

4.1.5. Die Konjugationseffizienz der Donoren und eines Transkonjuganten zeigte große Unterschiede

Die Anzahl der Transkonjuganten und die der Donoren wurde bestimmt und die Konjugationseffizienz berechnet. Diese Berechnung wird aus dem Verhältnis von Transkonjuganten zu Donoren ermittelt (Beaber et al., 2004). Für den Transfer von dem Donor NU14 Cm F' (HPI und *pks*-Insel) auf den Rezipienten wurde eine Übertragungsrate von $2,94 \times 10^{-7}$ cfu/ml errechnet. Die Konjugationseffizienz des Transfers von dem Stamm 536 Cm F' (HPI, *pks*- und *serU*-Insel) auf MG1655 Nal beträgt $3,85 \times 10^{-7}$ cfu/ml. Die Ähnlichkeit der beiden errechneten Werte ist nicht sehr verwunderlich, da die Übertragung von genomischen Material mit einem F'-Plasmid unabhängig von dem genetischen Inhalt ist (Griffiths et al., 1999). Als Vergleich zu der PAI-Übertragung wurde auch die Konjugationseffizienz des F'-Plasmid-Transfers ohne Berücksichtigung auf dessen genetischen Inhalt berechnet. Hierfür wurden nach der Konjugation nur Streptomycin-(Rezipient) und Tetracyclin-resistente (F'-Plasmid) Isolate gezählt. Die errechnete Übertragungsrate ist mit einem Wert von $5,24 \times 10^{-4}$ cfu/ml deutlich höher als die der PAI-Transfers. Von 120 überprüften Transkonjuganten wurde kein Isolat gefunden, welches eine Chloramphenicol-Resistenz und somit eine HPI hat. Da die Übertragung des reinen Plasmids um drei log-Stufen höher ist als die der PAIs, müssten weitaus mehr Isolate überprüft werden, um einen Chloramphenicol-resistenten Keim zu finden. Die hier gemessene Konjugationseffizienz stimmt auch mit einem bereits publizierten Wert überein. Es wurde berichtet, dass die Übertragungsrate eines F-Plasmids von einem Donor auf einen Rezipienten bei 1×10^{-4} cfu/ml liegt (Adelberg und Pittard, 1965). Um zu überprüfen, ob und wie effizient ein Tetracyclin-resistenter Transkonjugant sein neu erworbenes F'-Plasmid inklusive PAIs an einen weiteren Rezipienten weitergeben kann, wurde ein letzter Konjugationsversuch durchgeführt. Der Transkonjugant MG1655 HPI-Cm F' entstand durch die Konjugation NU14 Cm F' x MG1655 und wurde als Donor zur Übertragung auf den Stamm MG1655 Nal verwendet. Die errechnete Konjugationseffizienz ist in diesem Fall $4,14 \times 10^{-5}$ cfu/ml, welche eher der des reinen Plasmid-Transfers als der des PAI-Transfers ähnelt. Die hohe Übertragungsrate ist damit zu begründen, dass die PAIs bei einer vorhandenen Tetracyclin-Resistenz meist noch auf dem F'-Plasmid vorliegen und noch nicht in das Genom integriert wurden. Dadurch wurde in den meisten Fällen das Plasmid inklusive genomischen Inhalt direkt an den nächsten Rezipienten übertragen. Zudem wurde berichtet, dass die Frequenz des *Crossing-Over* nach dem erfolgten Transfer umso höher ist, je mehr homologe Regionen zwischen der DNA auf dem F'-Plasmid und der DNA des Rezipienten vorhanden sind

(Adelberg und Pittard, 1965). Da in diesem Versuch der *E. coli*-Stamm MG1655 als Donor und dessen Nalidixinsäure-resistente Mutante als Rezipient verwendet wurde, bestand dementsprechend eine höhere Homologie im Vergleich zu den vorherigen *E. coli*-Donoren NU14 und 536. Der Abbau des Plasmids im Transkonjuganten ohne vorherige Rekombination wäre daher geringer und der Einbau und die Detektion der PAIs höher.

Es wurde bereits ein ähnlicher Transfer gezeigt, in dem mit der Hilfe eines F'-Plasmids die HPI des Stammes NU14 auf den *E. coli*-Rezipienten AB1157 übertragen werden konnte. Im Vergleich war die Übertragung der HPI mit einer Konjugationseffizienz von 1×10^{-9} deutlich geringer (Schubert et al., 2009). Möglicherweise ist die Konjugationseffizienz abhängig von dem Rezipienten. Hierfür müsste geklärt werden, wie hoch die Homologien zwischen den Donoren und den Stämmen MG1655 und AB1157 sind. Bei dem erwähnten Transfer konnte auch per PCR gezeigt werden, dass auch einige hundert bp *upstream* der HPI durch das F'-Plasmid mit übertragen wurden (Schubert et al., 2009). Allerdings wurde in diesen Versuchen nicht das Ausmaß der insgesamt transferierten Donor-DNA analysiert. Zudem wurde nicht untersucht, ob weitere PAIs durch das F'-Plasmid mit übertragen wurden.

4.1.6. Die Menge an übertragener DNA war unabhängig von dem Donor und sehr variabel

Es wurde die Menge an transferierter Donor-DNA *in silico* ermittelt. Hierbei sind die Unterschiede sehr groß, aber vergleichbar zwischen dem Transfer von zwei und von drei Inseln und somit auch unabhängig von dem eingesetzten Donor. Die Spanne an übertragener DNA liegt zwischen ca. 62,5 kb und ca. 470 kb (Tab. 25). Die letztere Größe entspricht in etwa 10% des gesamten *E. coli*-Genoms. Zudem wird aber meist nicht die gesamte auf dem F'-Plasmid befindliche DNA in den Rezipienten durch homologe Rekombination integriert, da, wie bereits erwähnt, das *Crossing-Over* theoretisch an jeder Stelle, die zwischen Donor und Rezipient homolog ist, stattfinden kann (Griffiths et al., 1999). Addiert man noch die Länge des F'-Plasmids von 100 kb hinzu, ist die gesamte DNA die insgesamt übertragen wurde, enorm. Es wurde bereits von einem außergewöhnlich großem F-Plasmid (F14) berichtet, dessen DNA 10% des *E. coli*-Genoms repräsentiert (Pittard und Adelberg, 1964; Adelberg und Pittard, 1965).

Dass es sich bei den in dieser Arbeit durchgeführten Konjugationen um die Übertragung von einem Hfr-Stamm handelt, ist eher unwahrscheinlich (Baron et al., 1959). Zum einen wird bei

diesem Vorgang nur einzelsträngige DNA übertragen, die ohne Integration in den Rezipienten verloren geht, da sie sich nicht selbst replizieren kann und abgebaut wird (Griffiths et al., 1999). Somit hätte die extrachromosomale DNA nicht über mehrere Passagen über die Tetracyclin-Resistenz nachgewiesen werden können. Zum anderen würde bei einer Hfr-Übertragung das integrierte F-Plasmid zuletzt transferiert werden, was aber eher eine Seltenheit ist, da die Konjugation meistens bereits vorher abbricht (Griffiths et al., 1999). Bei der *in silico*-Auswertung des Tetracyclin-resistenten Transkonjuganten wurde das gesamte F'-Plasmid nachgewiesen, was wiederum gegen einen Transfer ausgehend von einem Hfr-Stamm spricht.

Bei einigen Transkonjuganten wurde mittels PCR festgestellt, dass die *pks*-Insel nicht zusammen mit den anderen Inseln übertragen wurde, obwohl sie im Donor vorhanden ist. Der Nachweis bei diesen Stämmen wurde auch *in silico* nicht erbracht. Auch Teile dieser Insel wurden wegen den bereits genannten Homologie-Abweichungen nicht gefunden. Das Vorhandensein der zwei unbesetzten *asn* tRNAs zwischen der HPI und der *pks*-Insel scheint daran nicht beteiligt gewesen zu sein, da die erste tRNA ca. 8,3 kb *downstream* der HPI lokalisiert ist, die Integration der Donor-DNA in den beiden sequenzierten Transkonjuganten aber bereits nach ca. 2,5 kb bzw. 2,8 kb endet. Die Rekombinationsstelle zwischen der HPI und der *pks*-Insel scheint daher zufällig zu sein, da rein theoretisch ein *Crossing-Over* an jeder homologen Stelle stattfinden kann (Griffiths et al., 1999), was auch durch die unterschiedlichen Orte der eingebauten DNA in Tabelle 25 gezeigt wird. Es wurden keine Transkonjuganten identifiziert, die keine *serU*-Insel erhalten haben. Allerdings ist die Sequenzlänge der inter-PAI-Region zwischen dieser PAI und der HPI nur ca. 1 kb und bietet somit weniger Möglichkeiten für eine homologe Rekombination.

Die Hälfte aller Transkonjuganten hatte nach der Konjugation eine Tetracyclin-Resistenz, was für das Vorhandensein des F'-Plasmids sprach. Bei der anderen Hälfte wurde die DNA des Donors bereits in den Rezipienten integriert und das Plasmid war ohne Antibiotika-Druck verloren gegangen, was *in silico* bestätigt wurde. Es wurde untersucht, was mit dem F'-Plasmid nach dessen Übertragung geschieht und wie der Einbau der Donoren-DNA in den Rezipienten durch homologe Rekombination verläuft. Es wurden 10 Tetracyclin-resistente Transkonjuganten ohne Antibiotika-Druck inkubiert. Man konnte sehen, dass über mehrere Passagen die Resistenz und somit auch das F'-Plasmid verloren ging, die mit einer Chloramphenicol-Resistenzkassette markierte HPI aber weiterhin vorhanden war. Zu welchem Zeitpunkt eine Rekombination stattgefunden hatte, konnte mit diesem Versuch aber

nicht geklärt werden. Ein passagierter Stamm und seine resistente Variante wurden in die Sequenzierungen mit eingeschlossen. Die beiden Isolate zeigen dieselben Rekombinationsstellen. Dies deutet darauf hin, dass die Rekombination bereits vor der Sequenzierung der resistenten Variante stattgefunden hatte. Zudem wurde in beiden keine ausgetauschte Rezipienten-DNA mehr gefunden. Möglicherweise wurde diese auf dem F'-Plasmid durch die flankierenden IS-Elemente ausgeschnitten und anschließend enzymatisch verdaut (Deonier und Mirels, 1977).

4.1.7. Die Hexapeptid-Wiederholungen des Gens *yaiX* könnten für den Abbruch der *Contigs* downstream der *pks*-Insel verantwortlich sein

Alle Stellen, an denen ein Übergang von Donor- zu Rezipienten-DNA gefunden wurde, wurden in dem Genom des Rezipienten MG1655 *in silico* auf mobile Elemente überprüft. Da an diesen Stellen keine dieser Elemente gefunden wurden, bekräftigt dies die Aussage, dass die Rekombination zwischen F'-Plasmid und Rezipienten an jeder homologen Region stattfinden kann (Griffiths et al., 1999). Bei allen Isolaten, die eine *pks*-Insel erhielten, bricht das *Contig* nach ca. 12,2 kb *downstream* dieser Insel innerhalb des Glykosyl-Transferase-Gens *yaiP* ab. Bei den meisten Transkonjuganten ist an dieser Stelle aber kein Übergang von Donor zu Rezipienten DNA zu finden, sondern erst viel weiter im *downstream*-Bereich. Der Abbruch der *Contigs* ist 29 bp vor der Überlappung mit dem Gen *yaiX* zu finden. Da *yaiX* sechs Hexapeptid Wiederholungen besitzt, könnten in dieser Region Probleme bei der Generierung der *Contigs* mit den sequenzierten Rohdaten aufgetreten sein. Repetitive DNA-Elemente erzeugen sehr oft technische Schwierigkeiten bei der Generierung von *Alignments* aus den Rohdaten durch die verwendeten Assemblierungsprogramme (Treangen und Salzberg, 2012).

Die Homologie des Gens *yaiP* von MG1655 ist höher zu dem Glykosyl-Transferase-Gen von 536, welches *downstream* der PAI III lokalisiert ist, als zu dem Gen *yaiP* von 536, welches *downstream* der *pks*-Insel zu finden ist. Möglicherweise ist das Gen *downstream* der *pks*-Insel eine Duplikation von dem *downstream* der PAI III, welches sich über die Zeit verändert hat (Abb. 40). Solche Duplikationen entstehen z.B. durch ungleiches *Crossing-Over*. Es konnte gezeigt werden, dass diese Art des *Crossing-Overs* durch ein F'-Plasmid entstehen kann (Sukhodolets, 2006). Es wurde innerhalb der ECOR-Sammlung kein Zusammenhang zwischen dem Vorhandensein der Glykosyl-Transferasen und der HPI gefunden. Allerdings fehlen die beschriebenen Glykosyl-Transferasen in den Bakterienstämmen der

phylogenetischen Gruppe E. Es wurde durch eine Transposon-Mutagenese in dem *E. coli*-Stamm MG1655 festgestellt, dass das Gen *yaiP* nicht essentiell ist (Winterberg et al., 2005).

MG1655 besitzt an der Position im Genom, an der die Stämme NU14 und 536 *downstream* der *pks*-Insel das Gen *yaiP* tragen, keine Glykosyl-Transferase. Es konnte allerdings nicht gezeigt werden, dass in diesem Gen die Rekombination stattgefunden hat. Dies sollte auch nicht der Fall sein, da an dieser Stelle keine homologen Bereiche zwischen Donor und Rezipient vorhanden sind, diese aber für eine Rekombination nötig sind (Griffiths et al., 1999). Der genaue Ort der Integration konnte nicht geklärt werden, da in den sequenzierten Genomen der Transkonjuganten an dieser Stelle eine Lücke aufgrund der fehlenden *Contigs* auftritt.

4.1.8. Die Verteilung der ECOR65-PAIs könnte durch einen unvollständigen Transfer entstanden sein

Die Sequenzen des Stammes ECOR65, der laut Definition in die PAI-Gruppe 2a einzuordnen ist, zeigen in den erstellten Stammbäumen meist eine höhere Homologie zu der PAI-Gruppe 3 als zu seiner eigentlichen Gruppe (Abb. 25 und 29). Bezüglich der *pks*-Insel und den inter-PAI-Regionen war anzunehmen, dass dieser Stamm ursprünglich eine *serU*-Insel enthielt. Die Zugehörigkeit zu der PAI-Gruppe 3 wurde durch eine tiefere Analyse der inter-PAI-Regionen untersucht. Hier wurde vermutet, dass der Stamm ECOR65 die *serU*-Insel durch eine Deletion verloren hat, da diese genomische Region als "Rekombinations-Hotspot" beschrieben wurde (Touchon et al., 2009). Durch die Nachstellung des *en bloc*-Transfers mittels F'-Plasmid wurde allerdings auch gezeigt, dass nicht immer alle PAIs auf den Rezipienten übertragen werden. Dies würde auch das Fehlen der *serU*-Insel erklären. Allerdings erläutert keine dieser Hypothesen die Homologie-Abweichungen der HPI gegenüber den PAI-Gruppen 2a und 3 (Tab. 17). Es wurde überprüft, ob es sich bei den Inseln des Stammes ECOR65 um ursprüngliche ICE-Typen oder intermediäre Formen zwischen diesen und den *E. coli*-PAIs handelt. Die Stammbäume, die die unterschiedlichen HPI- (Abb. 43) und *pks*-Insel-Sequenzen (Abb. 45) darstellen, zeigen keine Verwandtschaft zwischen den ECOR65-PAIs und den ICE-Typ-PAIs. Somit kann ein fremder Ursprung der Inseln vorerst ausgeschlossen werden. Möglicherweise stellt der Stamm ECOR65 ein anderes Transfer-Ereignis dar, als die der untersuchten PAI-Gruppen. Eine weitere Möglichkeit wäre, dass dieser Stamm nur eine HPI besaß, da die Sequenz dieser Insel sich innerhalb der PAI-Gruppe 1 anordnet (Abb. 22). Durch einen "unvollständigen" *en bloc*-Transfer ausgehend von einem

PAI-Gruppe 3-Donor, in dem nur die *pks*-Insel und das umliegende *Backbone*-Genom, nicht aber eine neue HPI oder eine *serU*-Insel, übertragen wurde, könnte diese PAI-Verteilung zustande gekommen sein. In den hier durchgeführten *in vitro*-Transfers mit einem F'-Plasmid wurde gezeigt, dass nicht immer alle vorhandenen PAIs auf den Rezipienten übertragen werden.

4.2. Phylogenie und Transfer der mobilen Pathogenitätsinseln

Die Phylogenie der mobilen PAIs wurde anhand der ICE-Typ-HPI untersucht, die in der Familie *Enterobacteriaceae* verbreitet ist. Dieser Typ ist in verschiedene Regionen aufgeteilt und wurde in dieser Arbeit in den Arten *E. coli*, *K. pneumoniae* und *C. koseri* untersucht (Schubert et al., 2004a; Seth-Smith und Croucher, 2009). Für den Transfer der ICE-Typ-HPI, die innerhalb der Familie *Enterobacteriaceae* über die Gattungsgrenze hinaus übertragen wurde, wurden verschiedene Konditionen ausgetestet (Schubert et al., 2004a). Die erstellten Transkonjuganten wurden bezüglich des Integrationsortes und der Funktionalität der HPI untersucht.

4.2.1. Phylogenie der mobilen Pathogenitätsinseln

Für die phylogenetische Analyse wurden die Region X, die konservierte Kern-Region und die *pks*-Insel, die bei einigen Isolaten in der Region III zu finden ist, verglichen (Preisach, 2008; Schubert et al., 2004a; Putze et al., 2009). Die Isolate, die neben der Kern-Region sowohl eine Region X als auch eine *pks*-Insel in der Region III aufweisen, waren *K. pneumoniae* 1084 und *C. koseri* ATCC BAA-895. Die Stämme, die eine zusätzliche Region X, aber keine *pks*-Insel besitzen, waren *K. pneumoniae* HU86, HU145 und B-2495 und *E. coli* ECOR31. In den Isolaten *K. pneumoniae* HU245 und *E. coli* ED1a ist weder eine Region X noch eine *pks*-Insel zu finden.

Bei der Betrachtung der Stammbäume bezüglich der Kern-Region (Abb. 42), der Region X (Abb. 44) und der Region III (Abb. 45) ist zu erkennen, dass die Sequenzen sich größtenteils entsprechend der untersuchten ICE-Typen und weniger entsprechend der Bakterien-Arten anordnen. Die Sequenz-Homologie der *E. coli*-Stämme ECOR31 und ED1a (99,30%; 221 SNPs), bezogen auf die Kern-Region der HPI (Abb. 43), ist niedriger als die durchschnittliche Homologie der Stämme, die eine kurze Variante der HPI besitzen (99,53%; 149,7 SNPs). Allerdings handelt es sich bei diesen beiden ICE-HPIs auch nicht um denselben ICE-Typ, da

Unterschiede bezüglich des Vorhandenseins der Region X bestehen. Im Allgemeinen ist die Sequenz-Homologie der ICE-Kernregionen (99,22%; 247,2 SNPs) niedriger als die der kurzen HPI (99,53%; 149,7 SNPs).

Die höchste Homologie ist bei drei *K. pneumoniae*-Stämmen, die alle eine Region X, aber keine *pks*-Insel aufweisen, zu finden. Die Statistik der Kern-Region (Tab. 26) zeigt eine Homologie von 99,96% (13,3 SNPs) und die der Region X (Tab. 27) von 99,92% (25,7 SNPs). Da zwei dieser Stämme, HU86 und HU145, identische Sequenzen sowohl bezüglich der Kern-Region als auch der Region X haben, könnte es sich hier um dasselbe Isolat handeln. Die Sequenzen des dritten Isolats, *K. pneumoniae* B-2495, bilden einen Cluster mit den Sequenzen von HU86 und HU145 bei der Analyse der Kern-Region, allerdings nicht bei der der Region X. Obwohl bereits beschrieben wurde, dass das *K. pneumoniae*-ICE ähnlich dem *E. coli*-ICE ist, zeigen diese drei *K. pneumoniae*-Stämme geringe Gemeinsamkeiten mit *E. coli* ECOR31, der ebenfalls eine Region X und keine *pks*-Insel besitzt (Seth-Smith und Croucher, 2009).

Die Sequenz der Kern-Region von *K. pneumoniae* HU245, welcher wie der Stamm *E. coli* ED1a weder eine Region X noch eine *pks*-Insel besitzt, zeigt keine besondere Ähnlichkeit im Vergleich mit den anderen untersuchten Sequenzen, auch nicht mit denen der *K. pneumoniae*-Isolate. Der Stamm ED1a bildet einen Cluster mit *E. coli* ECOR31, was möglicherweise als Art-spezifisch anzusehen ist, obwohl es sich nicht um die gleichen Inseln handelt.

Die Sequenzen der Stämme *C. koseri* ATCC BAA-895 und *K. pneumoniae* 1084 sind bezüglich der Kern-Region (98,36%; 520 SNPs) und der Region X (99,87%; 43 SNPs) weniger homolog, bilden aber immer eigene Cluster. Allerdings sind die Sequenzen der Region III (Tab. 28) untereinander homologer (99,91%; 49 SNPs) als im Vergleich mit der *pks*-Insel der ECOR-Sammlung (99,89%; 58,1 SNPs). Diese Ergebnisse sprechen für eigentlich unterschiedliche ICE-Typen, die aber nach demselben Prinzip zusammengebaut wurden. Es ist einerseits bekannt, dass mobile Elemente (*pks*-Insel) in ein bereits bestehendes ICE (HPI) direkt integrieren können. Andererseits besteht auch die Möglichkeit, dass ein solches Element auch neben einem bestehenden ICE inserieren und nachfolgend über homologe Regionen rekombinieren kann. In diesem Fall verschmelzen diese zu einem einzigen, chimären ICE (Johnson und Grossman, 2015).

Der Vergleich mit den deletierten *E. coli*-HPIs zeigt, dass die Kernsequenzen der ICEs einen eigenen Cluster bilden und sich somit evolutionär bereits weiter von diesen entfernt haben.

Allerdings scheint die HPI des *K. pneumoniae*-Stammes HU245 näher mit der von ECOR72 verwandt zu sein. Die HPI dieses Stammes wird als intermediäre Form mit der ICE-Typ-HPI in Verbindung gebracht (Schubert et al., 2009).

Zusammengefasst kann gezeigt werden, dass mittels NGS detaillierte Untersuchungen der verschiedenen Regionen innerhalb der einzelnen ICE-Typ-HPIs durchgeführt werden können. Zudem können Vergleiche mit anderen PAIs, wie denen von ExPEC, gezogen werden.

4.2.2. Konjugationsbedingungen der mobilen Pathogenitätsinseln

Um den Transfer von mobilen PAIs innerhalb der Familie der *Enterobacteriaceae* zu untersuchen, wurden Donoren und Rezipienten von verschiedenen Bakterien-Arten ausgewählt, um die ICE-Typ-HPI Insel über die Gattungsgrenze hinaus zu übertragen.

Die Konditionen für den Transfer mit einem F'-Plasmid konnten nicht übernommen werden, da eine Konjugation so nicht erreicht werden konnte. Deswegen wurden unterschiedliche Konjugationsbedingungen getestet, um eine erfolgreiche Übertragung der HPI zu ermöglichen (siehe B.9.3.2. Konjugation von aktiv transferierbaren Pathogenitätsinseln). Es wurden verschiedene Donoren und Rezipienten eingesetzt um zu prüfen, welche sich am besten eignen. Hierfür wurden deren optimalen Wachstumstemperaturen gewählt (Carniel et al., 1996). Zudem wurde die Dauer der Inkubation von Donor und Rezipient variiert. Unter bestimmten Konditionen kann das ICE induziert werden, was zum Ausschneiden des Elements aus dem Chromosom führt und die Wahrscheinlichkeit einer Konjugation erhöht. Diese Aktivierungssignale sind teilweise von ICE zu ICE unterschiedlich, allerdings gibt es allgemeine Auslöser, die viele ICEs induzieren. Dazu gehören die SOS-Antwort, die Ausschüttung von Signalmolekülen, die Wachstumsphase und selektive Vorteile durch das ICE (Johnson und Grossman, 2015).

4.2.3. Der Donor *K. pneumoniae* HU86 Cm entwickelte eine spontane Antibiotika-Resistenz

Mit den *E. coli*-Donoren konnte unter allen ausgeübten Konjugationsbedingungen kein Transfer nachgewiesen werden. Nach den Konjugationen mit dem *K. pneumoniae*-Stamm HU86 war stets ein Bakterienwachstum auf den Selektiv-Agarplatten zu erkennen. Allerdings handelte es sich bei diesen Isolaten um den Donor, weshalb immer eine Überprüfung per

Massenspektrometrie vorgenommen wurde. Es ist bereits bekannt, dass *K. pneumoniae*-Stämme eine spontane Antibiotika-Resistenz gegenüber unterschiedlichen Stoffen in kurzer Zeit entwickeln können (Smith, 1976; George et al., 1995). Durch die Auslösung der SOS-Antwort konnte kein Transfer des ICE erzwungen werden. Allerdings ist das Ausschneiden der meisten ICEs selbst unter induzierenden Bedingungen nur in einer relativ kleinen Anzahl an Zellen zu beobachten (Johnson und Grossman, 2015).

4.2.4. Die Region X könnte eine Rolle für einen erfolgreichen Transfer spielen

Ein Transfer der ICE-Typ-HPI war nur mit dem Region X-positiven Donor *K. pneumoniae* HU86 Cm möglich. Es konnte jeweils ein Transkonjugant mit den Rezipienten *E. coli* MG1655 und *Y. enterocolitica* MRS40 erhalten werden. Die beiden erfolgreichen Konjugationen wurden nach demselben Protokoll durchgeführt. Möglicherweise war der Transfer nur durch das Vorhandensein der Region X im Donor HU86 Cm zustande gekommen. Die Region X besitzt hohe Homologien zu dem Virulenz-Plasmid pLVPK von *K. pneumoniae* (Preisach, 2008). Dieses Plasmid scheint an der Mobilisierung und dem Transfer von DNA beteiligt zu sein (Putze et al., 2009; Chen et al., 2004). Es wurde bereits gezeigt, dass das ICE des *K. pneumoniae*-Stammes NTUH-K2044 auf einen *K. pneumoniae*- und einen *E. coli*-Rezipienten übertragen werden kann. Dieses ICE enthält zwar keine Region X, aber die mittlere Region dieser Insel hat wiederum Ähnlichkeiten zu dem Plasmid pLVPK und besitzt unter anderem den *iroBCDN* Gen-Cluster (Lin et al., 2008; Seth-Smith und Croucher, 2009).

4.2.5. Die Yersiniabaktin-Synthese des Transkonjuganten *Y. enterocolitica* MRS40 ICE-Cm ist temperaturabhängig

Um die Funktionsfähigkeit der HPI nach dem Transfer über die Gattungsgrenze hinaus zu testen, wurden CAS-Agarplatten verwendet. CAS-Agar verfärbt sich bei der bakteriellen Produktion von Siderophoren, wie z.B. Yersiniabaktin (Martin et al., 2013). Hierfür wurden jeweils der Donor, der Rezipient und der Transkonjugant auf den Platten ausgestrichen und bei 30°C bzw. 37°C inkubiert (Abb. 47). Der Donor *K. pneumoniae* HU86 Cm produziert Siderophore bei beiden getesteten Temperaturen. Auch der Rezipient *E. coli* MG1655 und der Transkonjugant *E. coli* MG1655 ICE-Cm verfärbten den CAS-Agar bei beiden Temperaturen. Die Verfärbung durch den Rezipienten MG1655 ist nicht ungewöhnlich, da dieser und die

meisten anderen *E. coli*-Stämme das Siderophor Enterobaktin produzieren (Martin et al., 2013). Ob der Transkonjugant MG1655 ICE-Cm Yersiniabaktin sezerniert, konnte durch diesen Test nicht vollständig geklärt werden. Die *Y. enterocolitica*-Stämme MRS40 und MRS40 ICE-Cm zeigen bei der Inkubationstemperatur von 37°C keine Siderophor-Produktion. Bei 30°C ist eine Verfärbung des CAS-Agars durch den Transkonjuganten zu sehen. Dies spricht für eine Yersiniabaktin-Produktion, die nur bei dem Temperaturoptimum von *Y. enterocolitica* zu beobachten ist (Carniel et al., 1996). Im Allgemeinen ist davon auszugehen, dass die ICE-Typ-HPI nach der Übertragung auf andere bakterielle Gattungen noch funktionsfähig ist.

4.2.6. Die ICE-Typ-HPI des *K. pneumoniae*-Stammes HU86 Cm kann mehrfach in einen Rezipienten integriert werden

Die Integrationsorte der HPI in den Transkonjuganten wurden bestimmt. Der Stamm MRS40 ICE-Cm wurde in die Sequenzierungen mit eingeschlossen und *in silico* ausgewertet. Es konnte gezeigt werden, dass eine einfache Kopie des ICE übertragen und in die *asn* tRNA YEt054 entsprechend dem *Y. enterocolitica*-Stamm 8081 integriert wurde (Carniel et al., 1996; Buchrieser et al., 1998). Die HPI kann in dieser bakteriellen Gattung in allen drei vorkommenden *asn* tRNAs integrieren (Carniel, 1999). Der PAI-Integrationsort des Transkonjuganten MG1655 ICE-Cm wurde per PCR bestimmt, da dieser Stamm erst nach den durchgeführten Sequenzierungen erstellt werden konnte (Abb. 48). Hierfür wurden sechs unterschiedliche Einzelkolonien des Transkonjuganten verwendet, um eine nachträgliche Neuintegration des ICE ausschließen zu können. Die PCR-Ergebnisse zeigen für alle sechs Einzelkolonien, dass die ICE-Typ-HPI in drei von vier *asn* tRNAs integriert wurde (Abb. 49). Es wurden Konjugationsversuche gezeigt, bei denen der *K. pneumoniae*-Donor NTUH-K2044 ein ICE mehrfach auf einen *K. pneumoniae*-Rezipienten übertragen hat. Diese Inseln waren in unterschiedlichen *asn* tRNAs integriert (Lin et al., 2008; Seth-Smith und Croucher, 2009).

5. Schlussfolgerung und Übertragungsmodell

Ziel dieser Arbeit war es, den horizontalen Gentransfer (HGT) von Pathogenitätsinseln (PAIs) innerhalb der Familie *Enterobacteriaceae* unter Verwendung der Technologie *Next-Generation Sequencing* zu untersuchen. Dies wurde mitunter durch die erstmalige Genom-Sequenzierung der *Escherichia coli* Referenz (ECOR)-Sammlung erreicht. Der Fokus lag hierbei auf den *E. coli*-PAIs *High-Pathogenicity Island* (HPI), PAI III, *pks*- und *serU*-Insel. Es konnte gezeigt werden, dass die Inseln unabhängig voneinander in die Art *E. coli* übertragen wurden und anschließend unterschiedliche Kombinationen der drei benachbarten PAIs HPI, *pks*- und *serU*-Insel *en bloc* innerhalb der Art *E. coli* mittels HGT weitergegeben wurden. Bakterienstämme mit diesen Insel-Kombinationen wurden in PAI-Gruppen eingeteilt und zeigten für ihre Gruppe spezifische PAI-Subtypen. Die hohe Homologie der genomischen *Backbone*-Bereiche zwischen den Inseln bestätigt die Übertragung innerhalb der PAI-Gruppen. Durch *in vitro*-Konjugationsversuche mit einem F'-Plasmid konnten die Transfers der immobilen *E. coli*-PAIs *en bloc* nachgestellt werden. Anhand der Sequenzierung der entstandenen Transkonjuganten wurden neue Einblicke in die Übertragung großer DNA-Abschnitte erhalten. Obwohl die Integration von Donor-DNA in das Rezipienten-Genom durch *Crossing-Over* theoretisch an jeder homologen Stelle stattfinden kann, wurde ein möglicher Integrationsort nahe des Glykosyl-Transferase-Gens *yaiP* identifiziert.

Die ICE-Typ-HPI, die Gene für einen eigenständigen Transfer enthält, wurde über die Gattungsgrenze hinaus in einen *E. coli*- und einen *Y. enterocolitica*-Stamm transferiert. Da sich diese Übertragung als sehr ineffizient erwies, mussten erst die geeigneten Konjugationsbedingungen getestet werden. Der einzige Donor, mit dem ein Transfer erfolgreich war, war der Region X-positive *K. pneumoniae*-Stamm HU86. Möglicherweise spielt diese Region bei der Übertragung eine Rolle. Die Auswertung der Transkonjuganten zeigte, dass die HPI nach dem Transfer in eine andere bakterielle Gattung noch funktionsfähig ist und das Siderophor Yersiniabaktin synthetisiert. Während der *Y. enterocolitica*-Rezipient eine ICE-Typ-HPI erhielt, wurden in den *E. coli*-Rezipient drei Kopien übertragen.

Zusammengefasst wurde gezeigt, dass die HPI über die Gattungsgrenze hinaus übertragen werden kann und noch funktionsfähig ist. Innerhalb der Art *E. coli* ist diese zusammen mit anderen PAIs *en bloc* übertragbar. Diese Art der Übertragung konnte sowohl *in vitro* als auch *in silico* bestätigt werden.

6. Ausblick

In dieser Arbeit wurden die Transfermechanismen von immobilen und mobilen PAIs nachgestellt und die Entwicklung der Pathogenität untersucht. Durch den passiven Transfer mit einem F'-Plasmid konnten Virulenzfaktoren auf apathogene *E. coli*-Stämme übertragen werden. Diese waren anschließend in der Lage die PAIs an weitere apathogene Stämme mit einer höheren Konjugationseffizienz weiter zu geben. Ob die evolutionäre Verbreitung der Inseln auch tatsächlich über ein F'-Plasmid oder einen Hfr-Stamm vollzogen wurde, konnte hier nicht geklärt werden. Es sollte in diesem Falle aber untersucht werden, ob Glykosyl-Transferasen (insbesondere YaiP) und deren Umgebung als Rekombinationsort für ein F'-Plasmid bevorzugt werden. Da der Transfer der mobilen PAIs nur durch einen Region X-positiven Stamm zustande kam, sollte diese Region auf die Auswirkung der Übertragung analysiert werden. Die Konjugation könnte nach einer erfolgreichen Deletion dieser Region wiederholt werden, um Unterschiede in dem Transfer zu verfolgen.

Durch die Technologie *Next-Generation Sequencing* ist es möglich geworden, binnen kurzer Zeit relativ kostengünstig neben pathogenen Isolaten und ganzen Stammsammlungen auch das menschliche Genom zu entschlüsseln und zu analysieren. Durch die stetige Erhöhung des Durchsatzes und der *Read*-Längen steigt die Qualität der Sequenzen und vereinfacht den Zusammenbau ganzer Genome. Diese Technologie wird auch dazu dienen, die Mechanismen der Übertragung von Virulenzfaktoren und Antibiotikaresistenzen besser zu verstehen. Dadurch könnte deren Weitergabe verhindert werden, was immer wichtiger in Zeiten von multiresistenten und hoch pathogenen Bakterien wird.

Zusammenfassung

Die vorliegende Arbeit befasste sich mit Untersuchungen des horizontalen Gentransfers (HGT) von Pathogenitätsinseln (PAIs) innerhalb der Familie *Enterobacteriaceae*. Während der Transfer von mobilen PAIs wie z.B. integrative und konjugative Elemente (ICEs) bereits etwas genauer untersucht wurde, sind die Transfermechanismen von immobilen PAIs weitgehend unbekannt. Da gerade innerhalb der Pathogruppe der extraintestinal pathogenen *Escherichia coli* (ExPEC), die für die meisten *E. coli*-Infektionen außerhalb des Magen-Darm-Traktes verantwortlich sind, immobile PAIs verbreitet sind und zwischen diesen Isolaten übertragen werden, wurden diese Inseln genauer untersucht.

Diese Arbeit beschäftigte sich vor allem mit der relativ jungen Technologie *Next-Generation Sequencing* (NGS), durch die eine Kombination aus *in silico* und *in vitro*-Untersuchungen durchgeführt werden konnte. Zuerst wurde die Phylogenie der Art *E. coli* anhand der repräsentativen *E. coli* Referenz (ECOR)-Sammlung durch die Anwendung mehrerer Multilocus-Sequenz-Typisierungen (MLST) analysiert. Es wurde gezeigt, dass durch die Verwendung der gesamten *Housekeeping*-Gene - anstatt deren Fragmente - hybride Genome identifiziert werden können. Im Weiteren wurde nachgewiesen, dass Stämme der ECOR-Sammlung anhand der Verteilung der benachbarten PAIs *High-Pathogenicity Island* (HPI), *pks*- und *serU*-Insel in sogenannte PAI-Gruppen eingeteilt werden können. Diese Isolate hatten sowohl für ihre Gruppe spezifische PAI-Subtypen als auch ein nahezu identisches *Backbone*-Genom zwischen diesen Inseln. Da gezeigt werden konnte, dass diese Stämme keinen klonalen Ursprung hatten, war dies der Nachweis für einen *en bloc*-Transfer der immobilen PAIs einschließlich der angrenzenden *Backbone*-Bereiche.

Anschließend wurde der *in silico* postulierte *en bloc*-Transfer *in vitro* nachgestellt. Für die Konjugationen wurden zwei *E. coli*-Donoren aus unterschiedlichen PAI-Gruppen erstellt, deren HPIs mit einer Antibiotika-Resistenzkassette markiert wurden, um den Transfer der Inseln zu verfolgen. Mit Hilfe eines F'-Plasmids wurden die immobilen PAIs *en bloc* übertragen und die *in silico*-Hypothese *in vitro* bestätigt. Durch die Sequenzierung der Transkonjuganten wurde gezeigt, dass der Transfer der einzelnen Inseln nach einem "Alles oder nichts" Prinzip abläuft und dass die Menge an übertragener DNA sehr variabel war. Obwohl die Rekombinationsorte zwischen Donoren- und Rezipienten-DNA größtenteils zufällig waren, konnte eine möglicherweise bevorzugte Integrationsstelle neben der Glykosyl-Transferase YaiP identifiziert werden.

Außerdem wurde die mobile ICE-Typ-HPI innerhalb der Familie *Enterobacteriaceae* über die Gattungsgrenze hinaus transferiert. Es wurde gezeigt, dass diese Insel mehrfach im Genom des Rezipienten integrieren kann und anschließend noch funktionsfähig ist. Die Region X der ICE-Typ-HPI könnte eine entscheidende Rolle bezüglich der Übertragung spielen. Die *in silico*-Untersuchungen der unterschiedlichen ICE-Regionen zeigten ebenfalls, dass spezifische ICE-Typ-HPIs in verschiedenen Vertretern der *Enterobacteriaceae* zu finden sind.

Es konnte zusammengefasst in dieser Arbeit gezeigt werden, dass die ICE-Typ-HPI als erste der drei untersuchten Inseln in die Art *E. coli* gelangte und sich hier über eine intermediäre Form etablierte. Im weiteren Verlauf kamen unabhängig voneinander die *serU*- und die *pks*-Insel hinzu und wurden fortan *en bloc* innerhalb der Art *E. coli* übertragen. Dieser Transfer scheint noch weiterhin im Gange zu sein. Diese neuen Erkenntnisse könnten zukünftig helfen, durch ein besseres Verständnis dieser Transfermechanismen die Übertragung von Virulenzfaktoren und Antibiotika-Resistenzen einzudämmen.

Summary

The main topic of this study was the investigation of horizontal gene transfer (HGT) of pathogenicity islands (PAIs) in *Enterobacteriaceae*. While the transfer of mobile PAIs, e.g. integrative and conjugative elements (ICEs), has been studied more detailed, the transfer mechanisms of immobile PAIs are still not well understood. As immobile PAIs are mainly found and transmitted in extraintestinal pathogenic *Escherichia coli* (ExPEC), causing most *E. coli* infections outside the gastrointestinal tract, these PAIs were examined more closely.

This study deals mainly with the relatively young technology next-generation sequencing (NGS). Due to NGS, it was possible to combine *in silico* and *in vitro* approaches. First, the *E. coli* phylogeny was studied by the representative *E. coli* reference (ECOR) collection using multilocus sequence typing (MLST). It was demonstrated that hybrid genomes were detectable due to the utilization of the entire housekeeping genes instead of the respective fragments. In a next step, we showed that strains of the ECOR collection could be classified into so called PAI-groups due to the distribution of the neighboring PAIs, namely the high-pathogenicity island (HPI), the *pks*- and the *serU*-island. These isolates had both PAI-subtypes specific for their group and nearly identical backbone regions between these islands. As it was demonstrated that these strains had no clonal origin, this was the proof for an *en bloc* transfer of these immobile PAIs including the surrounding backbone genome.

Afterwards, the *in silico* postulated *en bloc* transfer was recreated *in vitro*. For the conjugations, two *E. coli* donors from different PAI-groups were constructed. Their HPIs were tagged with an antibiotic resistance cassette to track the transfer. The immobile PAIs were transmitted *en bloc* by an F'-plasmid and the *in silico* hypothesis was confirmed *in vitro*. Due to sequencing of the transconjugants, it was demonstrated that the transfer occurs in an "all or nothing" fashion and that the amount of transmitted DNA was highly variable. Although the recombination sites between the donors' and the recipients' DNA were mostly random, one potential integration site was identified next to the glycosyl transferase YaiP.

Furthermore, the mobile ICE-type HPI within the family *Enterobacteriaceae* was transferred crossing the species border. It was demonstrated that this island can integrate several times into the recipient's genome and is subsequently still functional. The region X of the ICE-type HPI could play a major role in the transmission. The *in silico* analyses of different ICE

regions also showed that specific ICE-type HPIs can be found in different representatives of the family *Enterobacteriaceae*.

Taken together, this study showed that the ICE-type HPI was the first of the three investigated islands to appear into the *E. coli* species and was established as an intermediate type. During the subsequent process, the *serU*- and the *pks*-island appeared independently and were further spread *en bloc* together with the HPI within the *E. coli* species. This transfer seems to be still ongoing. These new findings could help to better understand the transfer mechanisms and to reduce the spread of virulence factors and antibiotic resistances.

Anhang

Abbildungsverzeichnis

- Abb. 1: Schematische Darstellung des Achtman MLST-Schemas
- Abb. 2: Schematische Darstellung des Pasteur MLST-Schemas
- Abb. 3: Schematische Darstellung der Triplex-PCR
- Abb. 4: Schematische Darstellung des vertikalen und horizontalen Gentransfers
- Abb. 5: Schematische Darstellung der Übertragung von chromosomaler DNA durch ein F'-Plasmid
- Abb. 6: Schematische Darstellung der HPI
- Abb. 7: Schematische Darstellung der *pks*-Insel
- Abb. 8: Schematische Darstellung der *serU*-Insel
- Abb. 9: Schematische Darstellung der PAI III
- Abb. 10: Schematische Darstellung der "Sequenzierung durch Synthese"
- Abb. 11: Gel-Bild eines üN-Verdaus von gDNA
- Abb. 12: Verwendete Parameter für das Trimmen der *Reads*
- Abb. 13: Verwendete Parameter für den *Workflow* "Assemblierung an Referenz"
- Abb. 14: Verwendete Parameter für den *Workflow* "*de novo* Assemblierung"
- Abb. 15: DNA-Zugehörigkeit innerhalb der Transkonjuganten
- Abb. 16: Farb-Code der PAI-Gruppen und Symbole der phylogenetischen Gruppen
- Abb. 17: MLST-Stammbaum der Fragmente der sechs *Housekeeping*-Gene
- Abb. 18: MLST-Stammbaum der Fragmente der acht *Housekeeping*-Gene

- Abb. 19: MLST-Stammbaum der sechs vollständigen *Housekeeping*-Gene
- Abb. 20: Schematische Darstellung der sechs vollständigen *Housekeeping*-Gene
- Abb. 21: Stammbaum der sechs vollständigen *Housekeeping*-Gene aller HPI-positiven Stämme
- Abb. 22: Stammbaum der HPI
- Abb. 23: Stammbaum der sechs codierenden HPI-Gene
- Abb. 24: Graphische Darstellung der ECOR65-HPI
- Abb. 25: Stammbaum der *pks*-Insel
- Abb. 26: Stammbaum der *serU*-Insel
- Abb. 27: Stammbaum der sechs codierenden *serU*-Insel-Gene
- Abb. 28: Stammbaum der inter-PAI-Region zwischen der *serU*-Insel und der HPI
- Abb. 29: Stammbaum der inter-PAI-Region zwischen der HPI und der *pks*-Insel
- Abb. 30: Ausschnitt aus dem Genom des Stammes ECOR70
- Abb. 31: Stammbaum des PAI III-Teilstücks
- Abb. 32: Stammbaum des 10 kb langen *upstream*-Bereichs der PAI III
- Abb. 33: Stammbaum des 10 kb langen *upstream*-Bereichs der PAI III der phylogenetischen Gruppe B1 und des Stammes 536
- Abb. 34: Stammbaum des 10 kb langen *downstream*-Bereichs der PAI III
- Abb. 35: Stammbaum des 10 kb langen *downstream*-Bereichs der PAI III ohne ECOR70
- Abb. 36: Stammbaum des 10 kb langen *downstream*-Bereichs der PAI III der phylogenetischen Gruppe B1 und des Stammes 536
- Abb. 37: Untersuchung der um die HPI liegenden inter-PAI-Regionen des Stammes ECOR65
- Abb. 38: Untersuchung der ECOR65-Genomregionen direkt neben der *serU*- und der *pks*-Insel

- Abb. 39: Erstellung der Donoren und Konjugation mit einem F'-Plasmid
- Abb. 40: Darstellung der Transferasen im Genom des Stammes 536
- Abb. 41: Gelbild der untersuchten Transferasen
- Abb. 42: Stammbaum der konservierten Kern-Region der ICE-Typ-HPI
- Abb. 43: Stammbaum der konservierten Kern-Region aller untersuchten Stämme
- Abb. 44: Stammbaum der Region X der ICE-Typ-HPI
- Abb. 45: Stammbaum der *pks*-Insel aller untersuchten Stämme
- Abb. 46: Stammbaum der sechs codierenden *pks*-Insel-Gene aller untersuchten Stämme
- Abb. 47: Nachweis der Siderophor-Synthese auf CAS-Agar
- Abb. 48: Vereinfachtes Schema der PCR zur Untersuchung der Integration der ICE-Typ-HPI
- Abb. 49: Gelbild zur Untersuchung der Integration der ICE-Typ-HPI

Tabellenverzeichnis

- Tab. 1: Erläuterung des Phred-Qualitätswerts
- Tab. 2: Verwendete Geräte
- Tab. 3: Verwendete Labor-Kits
- Tab. 4: Verwendete Antibiotika
- Tab. 5: Verwendete Antibiotika-Plättchen
- Tab. 6: Verwendete Stämme für die Erstellung der phylogenetischen Stammbäume
- Tab. 7: Verwendete Stämme für den Transfer der immobilen PAIs
- Tab. 8: Verwendete Stämme für den Transfer der mobilen PAIs

Tab. 9:	Verwendete Plasmide
Tab. 10:	Verwendete Oligonukleotide
Tab. 11:	Zusammensetzung einer einzelnen PCR
Tab. 12:	Ablauf einer PCR
Tab. 13:	PCR-Ergebnisse zur phylogenetischen Untersuchung der immobilisierten PAIs
Tab. 14:	Prozentuale Verteilung der PAIs
Tab. 15:	Prozentuale Verteilung der PAI III
Tab. 16:	Statistische Auswertung der sechs vollständigen <i>Housekeeping</i> -Gene
Tab. 17:	Statistische Auswertung der HPI
Tab. 18:	Statistische Auswertung der <i>pks</i> -Insel
Tab. 19:	Statistische Auswertung der <i>serU</i> -Insel
Tab. 20:	Statistische Auswertung der inter-PAI-Region zwischen der <i>serU</i> -Insel und der HPI
Tab. 21:	Statistische Auswertung der inter-PAI-Region zwischen der HPI und der <i>pks</i> -Insel
Tab. 22:	Statistische Auswertung des PAI III-Teilstücks
Tab. 23:	Statistische Auswertung des 10 kb langen <i>upstream</i> -Bereichs der PAI III
Tab. 24:	Statistische Auswertung des 10 kb langen <i>downstream</i> -Bereichs der PAI III
Tab. 25:	Auswertung der Transkonjuganten
Tab. 26:	Statistische Auswertung der Kern-Region der ICE-Typ-HPI
Tab. 27:	Statistische Auswertung der Region X
Tab. 28:	Statistische Auswertung der <i>pks</i> -Insel aller untersuchten Stämme

Literaturverzeichnis

Teile dieser Arbeit wurden zur Publikation bei PLOS ONE eingereicht und die Genome der erstmalig sequenzierten ECOR-Sammlung auf NCBI hinterlegt:

Messerer, M., Fischer, W., Schubert, S. 2016 Investigation of Horizontal Gene Transfer of Pathogenicity Islands in *Escherichia coli* using Next-Generation Sequencing

Adelberg,E.A. and Pittard,J. (1965). Chromosome Transfer In Bacteriel Conjugation. Bacteriol. Rev. 29, 161-172.

Arun-Chinnappa,K.S. and McCurdy,D.W. (2015). De novo assembly of a genome-wide transcriptome map of *Vicia faba* (L.) for transfer cell research. Front Plant Sci. 6, 217.

Avery,O.T., MacLeod,C.M., and McCarty,M. (1979). Studies on the chemical nature of the substance inducing transformation of pneumococcal types. Inductions of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III. J. Exp. Med. 149, 297-326.

Bach,S., de,A.A., and Carniel,E. (2000). The Yersinia high-pathogenicity island is present in different members of the family Enterobacteriaceae. FEMS Microbiol. Lett. 183, 289-294.

Barasch,J. and Mori,K. (2004). Cell biology: iron thievery. Nature 432, 811-813.

Baron,L.S., Carey,W.F., and Spilman,W.M. (1959). Characteristics Of A High Frequency Of Recombination (HFR) Strain Of Salmonella Typhosa Compatible With Salmonella, Shigella, And Escherichia Species. Proc. Natl. Acad. Sci. U. S. A 45, 1752-1757.

Bauer,R.J., Zhang,L., Foxman,B., Siitonen,A., Jantunen,M.E., Saxen,H., and Marrs,C.F. (2002). Molecular epidemiology of 3 putative virulence genes for *Escherichia coli* urinary tract infection-usp, iha, and iroN(*E. coli*). J. Infect. Dis. 185, 1521-1524.

Baumler,A.J., Norris,T.L., Lasco,T., Voight,W., Reissbrodt,R., Rabsch,W., and Heffron,F. (1998). IroN, a novel outer membrane siderophore receptor characteristic of *Salmonella enterica*. J. Bacteriol. 180, 1446-1453.

Beaber,J.W., Hochhut,B., and Waldor,M.K. (2004). SOS response promotes horizontal dissemination of antibiotic resistance genes. Nature 427, 72-74.

Benedek,O. and Schubert,S. (2007). Mobility of the Yersinia High-Pathogenicity Island (HPI): transfer mechanisms of pathogenicity islands (PAIS) revisited (a review). Acta Microbiol. Immunol. Hung. 54, 89-105.

Bergthorsson,U. and Ochman,H. (1998). Distribution of chromosome length variation in natural isolates of *Escherichia coli*. Mol. Biol. Evol. 15, 6-16.

Bidet,P., Bonacorsi,S., Clermont,O., De,M.C., Brahimi,N., and Bingen,E. (2005). Multiple insertional events, restricted by the genetic background, have led to acquisition of pathogenicity island IIIJ96-like domains among *Escherichia coli* strains of different clinical origins. *Infect. Immun.* 73, 4081-4087.

Blattner,F.R., Plunkett,G., III, Bloch,C.A., Perna,N.T., Burland,V., Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F., Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J., Mau,B., and Shao,Y. (1997). The complete genome sequence of *Escherichia coli* K-12. *Science* 277, 1453-1462.

Bouchet,V., Huot,H., and Goldstein,R. (2008). Molecular genetic basis of ribotyping. *Clin. Microbiol. Rev.* 21, 262-73, table.

Bower,J.M., Eto,D.S., and Mulvey,M.A. (2005). Covert operations of uropathogenic *Escherichia coli* within the urinary tract. *Traffic.* 6, 18-31.

Buchrieser,C., Brosch,R., Bach,S., Guiyoule,A., and Carniel,E. (1998). The high-pathogenicity island of *Yersinia pseudotuberculosis* can be inserted into any of the three chromosomal *asn* tRNA genes. *Mol. Microbiol.* 30, 965-978.

Carniel,E. (1999). The *Yersinia* high-pathogenicity island. *Int. Microbiol.* 2, 161-167.

Carniel,E. (2001). The *Yersinia* high-pathogenicity island: an iron-uptake island. *Microbes. Infect.* 3, 561-569.

Carniel,E., Guilvout,I., and Prentice,M. (1996). Characterization of a large chromosomal "high-pathogenicity island" in biotype 1B *Yersinia enterocolitica*. *J. Bacteriol.* 178, 6743-6751.

Chaudhuri,R.R. and Henderson,I.R. (2012). The evolution of the *Escherichia coli* phylogeny. *Infect. Genet. Evol.* 12, 214-226.

Chen,Y.T., Chang,H.Y., Lai,Y.C., Pan,C.C., Tsai,S.F., and Peng,H.L. (2004). Sequencing and analysis of the large virulence plasmid pLVPK of *Klebsiella pneumoniae* CG43. *Gene* 337, 189-198.

Cirl,C., Wieser,A., Yadav,M., Duerr,S., Schubert,S., Fischer,H., Stappert,D., Wantia,N., Rodriguez,N., Wagner,H., Svanborg,C., and Miethke,T. (2008). Subversion of Toll-like receptor signaling by a unique family of bacterial Toll/interleukin-1 receptor domain-containing proteins. *Nat. Med.* 14, 399-406.

Clermont,O., Bonacorsi,S., and Bingen,E. (2000). Rapid and simple determination of the *Escherichia coli* phylogenetic group. *Appl. Environ. Microbiol.* 66, 4555-4558.

Clermont,O., Bonacorsi,S., and Bingen,E. (2001). The *Yersinia* high-pathogenicity island is highly predominant in virulence-associated phylogenetic groups of *Escherichia coli*. *FEMS Microbiol. Lett.* 196, 153-157.

Crosa,J.H. (1989). Genetics and molecular biology of siderophore-mediated iron transport in bacteria. *Microbiol. Rev.* 53, 517-530.

Datsenko,K.A. and Wanner,B.L. (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. U. S. A* 97, 6640-6645.

Denamur,E., Lecoindre,G., Darlu,P., Tenaillon,O., Acquaviva,C., Sayada,C., Sunjevaric,I., Rothstein,R., Elion,J., Taddei,F., Radman,M., and Matic,I. (2000). Evolutionary implications of the frequent horizontal transfer of mismatch repair genes. *Cell* 103, 711-721.

Deonier,R.C. and Mirels,L. (1977). Excision of F plasmid sequences by recombination at directly repeated insertion sequence 2 elements: involvement of recA. *Proc. Natl. Acad. Sci. U. S. A* 74, 3965-3969.

Dobrindt,U., Blum-Oehler,G., Nagy,G., Schneider,G., Johann,A., Gottschalk,G., and Hacker,J. (2002). Genetic structure and distribution of four pathogenicity islands (PAI I(536) to PAI IV(536)) of uropathogenic *Escherichia coli* strain 536. *Infect. Immun.* 70, 6365-6372.

Dobrindt,U., Chowdary,M.G., Krumbholz,G., and Hacker,J. (2010). Genome dynamics and its impact on evolution of *Escherichia coli*. *Med. Microbiol. Immunol.* 199, 145-154.

Dobrindt,U. and Hacker,J. (2001). Whole genome plasticity in pathogenic bacteria. *Curr. Opin. Microbiol.* 4, 550-557.

Dobrindt,U., Hochhut,B., Hentschel,U., and Hacker,J. (2004). Genomic islands in pathogenic and environmental microorganisms. *Nat. Rev. Microbiol.* 2, 414-424.

Enright,M.C. and Spratt,B.G. (1999). Multilocus sequence typing. *Trends Microbiol.* 7, 482-487.

Escherich,T. (1989). The intestinal bacteria of the neonate and breast-fed infant. 1885. *Rev. Infect. Dis.* 11, 352-356.

Escobar-Paramo,P., Clermont,O., Blanc-Potard,A.B., Bui,H., Le,B.C., and Denamur,E. (2004). A specific genetic background is required for acquisition and expression of virulence factors in *Escherichia coli*. *Mol. Biol. Evol.* 21, 1085-1094.

Ewing,B. and Green,P. (1998). Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* 8, 186-194.

Ewing,B., Hillier,L., Wendl,M.C., and Green,P. (1998). Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* 8, 175-185.

Frelinger,J.A. (2015). Big Data, Big Opportunities, and Big Challenges. *J. Investig. Dermatol. Symp. Proc.* 17, 33-35.

Frost,L.S., Ippen-Ihler,K., and Skurray,R.A. (1994). Analysis of the sequence and gene products of the transfer region of the F sex factor. *Microbiol. Rev.* 58, 162-210.

George,A.M., Hall,R.M., and Stokes,H.W. (1995). Multidrug resistance in *Klebsiella pneumoniae*: a novel gene, ramA, confers a multidrug resistance phenotype in *Escherichia coli*. *Microbiology* 141 (Pt 8), 1909-1920.

Griffith,F. (1928). The Significance of Pneumococcal Types. *J. Hyg. (Lond)* 27, 113-159.

Griffiths,A., Gelbart,W., Miller,J., and Lewontin,R. (1999). *Modern Genetic Analysis*.

Grimont,F. and Grimont,P.A. (1986). Ribosomal ribonucleic acid gene restriction patterns as potential taxonomic tools. *Ann. Inst. Pasteur Microbiol.* 137B, 165-175.

Groisman,E.A. and Ochman,H. (1996). Pathogenicity islands: bacterial evolution in quantum leaps. *Cell* 87, 791-794.

Hacker,J., Hochhut,B., Middendorf,B., Schneider,G., Buchrieser,C., Gottschalk,G., and Dobrindt,U. (2004). Pathogenomics of mobile genetic elements of toxigenic bacteria. *Int. J. Med. Microbiol.* 293, 453-461.

Hacker,J. and Kaper,J.B. (2000). Pathogenicity islands and the evolution of microbes. *Annu. Rev. Microbiol.* 54, 641-679.

Hacker,J., Ott,M., Blum,G., Marre,R., Heesemann,J., Tschape,H., and Goebel,W. (1992). Genetics of *Escherichia coli* uropathogenicity: analysis of the O6:K15:H31 isolate 536. *Zentralbl. Bakteriologie* 276, 165-175.

Hakenbeck,R., Balmelle,N., Weber,B., Gardes,C., Keck,W., and de,S.A. (2001). Mosaic genes and mosaic chromosomes: intra- and interspecies genomic variation of *Streptococcus pneumoniae*. *Infect. Immun.* 69, 2477-2486.

Hancock,V., Ferrieres,L., and Klemm,P. (2008). The ferric yersiniabactin uptake receptor FyuA is required for efficient biofilm formation by urinary tract infectious *Escherichia coli* in human urine. *Microbiology* 154, 167-175.

Hantke,K., Nicholson,G., Rabsch,W., and Winkelmann,G. (2003). Salmochelins, siderophores of *Salmonella enterica* and uropathogenic *Escherichia coli* strains, are recognized by the outer membrane receptor IroN. *Proc. Natl. Acad. Sci. U. S. A* 100, 3677-3682.

Herzer,P.J., Inouye,S., Inouye,M., and Whittam,T.S. (1990). Phylogenetic distribution of branched RNA-linked multicopy single-stranded DNA among natural isolates of *Escherichia coli*. *J. Bacteriol.* 172, 6175-6181.

Homburg,S., Oswald,E., Hacker,J., and Dobrindt,U. (2007). Expression analysis of the colibactin gene cluster coding for a novel polyketide in *Escherichia coli*. *FEMS Microbiol. Lett.* 275, 255-262.

Jauregui,F., Landraud,L., Passet,V., Diancourt,L., Frapy,E., Guigon,G., Carbonnelle,E., Lortholary,O., Clermont,O., Denamur,E., Picard,B., Nassif,X., and Brisse,S. (2008). Phylogenetic and genomic diversity of human bacteremic *Escherichia coli* strains. *BMC Genomics* 9, 560.

Johnson,C.M. and Grossman,A.D. (2015). Integrative and Conjugative Elements (ICEs): What They Do and How They Work. *Annu. Rev. Genet.* 49, 577-601.

Johnson,J.R. (1991). Virulence factors in *Escherichia coli* urinary tract infection. *Clin. Microbiol. Rev.* 4, 80-128.

Johnson,J.R., Brown,J.J., Carlino,U.B., and Russo,T.A. (1998). Colonization with and acquisition of uropathogenic *Escherichia coli* as revealed by polymerase chain reaction-based detection. *J. Infect. Dis.* 177, 1120-1124.

Johnson,J.R., Delavari,P., Kuskowski,M., and Stell,A.L. (2001a). Phylogenetic distribution of extraintestinal virulence-associated traits in *Escherichia coli*. *J. Infect. Dis.* 183, 78-88.

- Johnson,J.R. and Russo,T.A. (2002). Extraintestinal pathogenic *Escherichia coli*: "the other bad *E coli*". *J. Lab Clin. Med.* *139*, 155-162.
- Johnson,J.R. and Stell,A.L. (2000). Extended virulence genotypes of *Escherichia coli* strains from patients with urosepsis in relation to phylogeny and host compromise. *J. Infect. Dis.* *181*, 261-272.
- Johnson,J.R., Weissman,S.J., Stell,A.L., Trintchina,E., Dykhuizen,D.E., and Sokurenko,E.V. (2001b). Clonal and pathotypic analysis of archetypal *Escherichia coli* cystitis isolate NU14. *J. Infect. Dis.* *184*, 1556-1565.
- Kauffmann,F. (1947). The serology of the coli group. *J. Immunol.* *57*, 71-100.
- Koczura,R. and Kaznowski,A. (2003). Occurrence of the *Yersinia* high-pathogenicity island and iron uptake systems in clinical isolates of *Klebsiella pneumoniae*. *Microb. Pathog.* *35*, 197-202.
- Larsen,M.V., Cosentino,S., Rasmussen,S., Friis,C., Hasman,H., Marvig,R.L., Jelsbak,L., Sicheritz-Ponten,T., Ussery,D.W., Aarestrup,F.M., and Lund,O. (2012). Multilocus sequence typing of total-genome-sequenced bacteria. *J. Clin. Microbiol.* *50*, 1355-1361.
- Lawlor,M.S., O'connor,C., and Miller,V.L. (2007). *Yersiniabactin* is a virulence factor for *Klebsiella pneumoniae* during pulmonary infection. *Infect. Immun.* *75*, 1463-1472.
- Lawrence,J.G. (1999). Gene transfer, speciation, and the evolution of bacterial genomes. *Curr. Opin. Microbiol.* *2*, 519-523.
- Le,G.T., Clermont,O., Gouriou,S., Picard,B., Nassif,X., Denamur,E., and Tenaillon,O. (2007). Extraintestinal virulence is a coincidental by-product of commensalism in B2 phylogenetic group *Escherichia coli* strains. *Mol. Biol. Evol.* *24*, 2373-2384.
- Lecointre,G., Rachdi,L., Darlu,P., and Denamur,E. (1998). *Escherichia coli* molecular phylogeny using the incongruence length difference test. *Mol. Biol. Evol.* *15*, 1685-1695.
- Lederberg,J. and Tatum,E.L. (1953). Sex in bacteria; genetic studies, 1945-1952. *Science* *118*, 169-175.
- Lee,J., Izzah,N.K., Choi,B.S., Joh,H.J., Lee,S.C., Perumal,S., Seo,J., Ahn,K., Jo,E.J., Choi,G.J., Nou,I.S., Yu,Y., and Yang,T.J. (2015). Genotyping-by-sequencing map permits identification of clubroot resistance QTLs and revision of the reference genome assembly in cabbage (*Brassica oleracea* L.). *DNA Res.*
- Lin,A.C., Liao,T.L., Lin,Y.C., Lai,Y.C., Lu,M.C., and Chen,Y.T. (2012). Complete genome sequence of *Klebsiella pneumoniae* 1084, a hypermucoviscosity-negative K1 clinical strain. *J. Bacteriol.* *194*, 6316.
- Lin,T.L., Lee,C.Z., Hsieh,P.F., Tsai,S.F., and Wang,J.T. (2008). Characterization of integrative and conjugative element ICEKp1-associated genomic heterogeneity in a *Klebsiella pneumoniae* strain isolated from a primary liver abscess. *J. Bacteriol.* *190*, 515-526.
- Liu,L., Li,Y., Li,S., Hu,N., He,Y., Pong,R., Lin,D., Lu,L., and Law,M. (2012). Comparison of next-generation sequencing systems. *J. Biomed. Biotechnol.* *2012*, 251364.

Luthra,R., Chen,H., Roy-Chowdhuri,S., and Singh,R.R. (2015). Next-Generation Sequencing in Clinical Molecular Diagnostics of Cancer: Advantages and Challenges. *Cancers. (Basel)* 7, 2023-2036.

Lyubetsky,V., Gershgorin,R., Seliverstov,A., and Gorbunov,K. (2016). Algorithms for reconstruction of chromosomal structures. *BMC. Bioinformatics.* 17, 40.

Maiden,M.C., Bygraves,J.A., Feil,E., Morelli,G., Russell,J.E., Urwin,R., Zhang,Q., Zhou,J., Zurth,K., Caugant,D.A., Feavers,I.M., Achtman,M., and Spratt,B.G. (1998). Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc. Natl. Acad. Sci. U. S. A* 95, 3140-3145.

Manzoor,U., Shahid,S., and Zafar,B. (2015). A comparative analysis of multiple sequence alignments for biological data. *Biomed. Mater. Eng* 26 *Suppl 1*, S1781-S1789.

Martin,P., Marcq,I., Magistro,G., Penary,M., Garcie,C., Payros,D., Boury,M., Olier,M., Nougayrede,J.P., Audebert,M., Chalut,C., Schubert,S., and Oswald,E. (2013). Interplay between siderophores and colibactin genotoxin biosynthetic pathways in *Escherichia coli*. *PLoS. Pathog.* 9, e1003437.

Metzker,M.L. (2010). Sequencing technologies - the next generation. *Nat. Rev. Genet.* 11, 31-46.

Meyer,C., Hoffmann,C., Haas,R., and Schubert,S. (2015). The role of the galU gene of uropathogenic *Escherichia coli* in modulating macrophage TNF-alpha response. *Int. J. Med. Microbiol.* 305, 893-901.

Milkman,R. (1973). Electrophoretic variation in *Escherichia coli* from natural sources. *Science* 182, 1024-1026.

Murray,A.C., Kuskowski,M.A., and Johnson,J.R. (2004). Virulence factors predict *Escherichia coli* colonization patterns among human and animal household members. *Ann. Intern. Med.* 140, 848-849.

Muzzey,D., Evans,E.A., and Lieber,C. (2015). Understanding the Basics of NGS: From Mechanism to Variant Calling. *Curr. Genet. Med. Rep.* 3, 158-165.

Nakai,H., Hagihara,M., Kato,H., Hirai,J., Nishiyama,N., Koizumi,Y., Sakanashi,D., Suematsu,H., Yamagishi,Y., and Mikamo,H. (2016). Prevalence and risk factors of infections caused by extended-spectrum beta-lactamase (ESBL)-producing Enterobacteriaceae. *J. Infect. Chemother.*

Negre,V.L., Bonacorsi,S., Schubert,S., Bidet,P., Nassif,X., and Bingen,E. (2004). The siderophore receptor IroN, but not the high-pathogenicity island or the hemin receptor ChuA, contributes to the bacteremic step of *Escherichia coli* neonatal meningitis. *Infect. Immun.* 72, 1216-1220.

Neilands,J.B. (1976). Siderophores: diverse roles in microbial and human physiology. *Ciba Found. Symp.* 107-124.

Norenberg,D., Wieser,A., Magistro,G., Hoffmann,C., Meyer,C., Messerer,M., and Schubert,S. (2013). Molecular analysis of a novel Toll/interleukin-1 receptor (TIR)-domain containing

virulence protein of *Y. pseudotuberculosis* among Far East scarlet-like fever serotype I strains. *Int. J. Med. Microbiol.* 303, 583-594.

Nougayrede,J.P., Homburg,S., Taieb,F., Boury,M., Brzuszkiewicz,E., Gottschalk,G., Buchrieser,C., Hacker,J., Dobrindt,U., and Oswald,E. (2006). *Escherichia coli* induces DNA double-strand breaks in eukaryotic cells. *Science* 313, 848-851.

Nowrouzian,F.L., Wold,A.E., and Adlerberth,I. (2005). *Escherichia coli* strains belonging to phylogenetic group B2 have superior capacity to persist in the intestinal microflora of infants. *J. Infect. Dis.* 191, 1078-1083.

Ochman,H. and Bergthorsson,U. (1998). Rates and patterns of chromosome evolution in enteric bacteria. *Curr. Opin. Microbiol.* 1, 580-583.

Ochman,H., Lawrence,J.G., and Groisman,E.A. (2000). Lateral gene transfer and the nature of bacterial innovation. *Nature* 405, 299-304.

Ochman,H. and Selander,R.K. (1984). Standard reference strains of *Escherichia coli* from natural populations. *J. Bacteriol.* 157, 690-693.

Orskov,F. and Orskov,I. (1992). *Escherichia coli* serotyping and disease in man and animals. *Can. J. Microbiol.* 38, 699-704.

Pelludat,C., Hogardt,M., and Heesemann,J. (2002). Transfer of the core region genes of the *Yersinia enterocolitica* WA-C serotype O:8 high-pathogenicity island to *Y. enterocolitica* MRS40, a strain with low levels of pathogenicity, confers a yersiniabactin biosynthesis phenotype and enhanced mouse virulence. *Infect. Immun.* 70, 1832-1841.

Picard,B., Garcia,J.S., Gouriou,S., Duriez,P., Brahimi,N., Bingen,E., Elion,J., and Denamur,E. (1999). The link between phylogeny and virulence in *Escherichia coli* extraintestinal infection. *Infect. Immun.* 67, 546-553.

Pittard,J. and Adelberg,E.A. (1964). Gene Transfer By F' Strains Of *Escherichia Coli* K-12. III. An Analysis Of The Recombination Events Occuring In The F' Male And In The Zygotes. *Genetics* 49, 995-1007.

Preisach, B. Charakterisierung einer horizontal transferierten Region im *Escherichia coli* Stamm ECOR31. 11-12-2008. LMU München.

Ref Type: Thesis/Dissertation

Putze,J., Hennequin,C., Nougayrede,J.P., Zhang,W., Homburg,S., Karch,H., Bringer,M.A., Fayolle,C., Carniel,E., Rabsch,W., Oelschlaeger,T.A., Oswald,E., Forestier,C., Hacker,J., and Dobrindt,U. (2009). Genetic structure and distribution of the colibactin genomic island among members of the family Enterobacteriaceae. *Infect. Immun.* 77, 4696-4703.

Rakin,A., Saken,E., Harmsen,D., and Heesemann,J. (1994). The pesticin receptor of *Yersinia enterocolitica*: a novel virulence factor with dual function. *Mol. Microbiol.* 13, 253-263.

Raymond,K.N., Dertz,E.A., and Kim,S.S. (2003). Enterobactin: an archetype for microbial iron transport. *Proc. Natl. Acad. Sci. U. S. A* 100, 3584-3588.

Russo,T.A., Carlino,U.B., Mong,A., and Jodush,S.T. (1999). Identification of genes in an extraintestinal isolate of *Escherichia coli* with increased expression after exposure to human urine. *Infect. Immun.* *67*, 5306-5314.

Russo,T.A., McFadden,C.D., Carlino-MacDonald,U.B., Beanan,J.M., Barnard,T.J., and Johnson,J.R. (2002). *IroN* functions as a siderophore receptor and is a urovirulence factor in an extraintestinal pathogenic isolate of *Escherichia coli*. *Infect. Immun.* *70*, 7156-7160.

Russo,T.A., McFadden,C.D., Carlino-MacDonald,U.B., Beanan,J.M., Olson,R., and Wilding,G.E. (2003). The Siderophore receptor *IroN* of extraintestinal pathogenic *Escherichia coli* is a potential vaccine candidate. *Infect. Immun.* *71*, 7164-7169.

Sahl,J.W., Matalaka,M.N., and Rasko,D.A. (2012). Phylomark, a tool to identify conserved phylogenetic markers from whole-genome alignments. *Appl. Environ. Microbiol.* *78*, 4884-4892.

Schmidt,H. and Hensel,M. (2004). Pathogenicity islands in bacterial pathogenesis. *Clin. Microbiol. Rev.* *17*, 14-56.

Schubert,S., Cuenca,S., Fischer,D., and Heesemann,J. (2000). High-pathogenicity island of *Yersinia pestis* in enterobacteriaceae isolated from blood cultures and urine samples: prevalence and functional expression. *J. Infect. Dis.* *182*, 1268-1271.

Schubert,S., Darlu,P., Clermont,O., Wieser,A., Magistro,G., Hoffmann,C., Weinert,K., Tenaillon,O., Matic,I., and Denamur,E. (2009). Role of intraspecies recombination in the spread of pathogenicity islands within the *Escherichia coli* species. *PLoS. Pathog.* *5*, e1000257.

Schubert,S., Dufke,S., Sorsa,J., and Heesemann,J. (2004a). A novel integrative and conjugative element (ICE) of *Escherichia coli*: the putative progenitor of the *Yersinia* high-pathogenicity island. *Mol. Microbiol.* *51*, 837-848.

Schubert,S., Norenberg,D., Clermont,O., Magistro,G., Wieser,A., Romann,E., Hoffmann,C., Weinert,K., and Denamur,E. (2010). Prevalence and phylogenetic history of the *TcpC* virulence determinant in *Escherichia coli*. *Int. J. Med. Microbiol.* *300*, 429-434.

Schubert,S., Picard,B., Gouriou,S., Heesemann,J., and Denamur,E. (2002). *Yersinia* high-pathogenicity island contributes to virulence in *Escherichia coli* causing extraintestinal infections. *Infect. Immun.* *70*, 5335-5337.

Schubert,S., Rakin,A., and Heesemann,J. (2004b). The *Yersinia* high-pathogenicity island (HPI): evolutionary and functional aspects. *Int. J. Med. Microbiol.* *294*, 83-94.

Schubert,S., Rakin,A., Karch,H., Carniel,E., and Heesemann,J. (1998). Prevalence of the "high-pathogenicity island" of *Yersinia* species among *Escherichia coli* strains that are pathogenic to humans. *Infect. Immun.* *66*, 480-485.

Seth-Smith,H. and Croucher,N.J. (2009). Genome watch: breaking the ICE. *Nat. Rev. Microbiol.* *7*, 328-329.

Siitonen,A. (1992). *Escherichia coli* in fecal flora of healthy adults: serotypes, P and type 1C fimbriae, non-P mannose-resistant adhesins, and hemolytic activity. *J. Infect. Dis.* *166*, 1058-1065.

Smith,H.W. (1976). Mutants of *Klebsiella pneumoniae* resistant to several antibiotics. *Nature* 259, 307-308.

Sorokin,A., Candelon,B., Guilloux,K., Galleron,N., Wackerow-Kouzova,N., Ehrlich,S.D., Bourguet,D., and Sanchis,V. (2006). Multiple-locus sequence typing analysis of *Bacillus cereus* and *Bacillus thuringiensis* reveals separate clustering and a distinct population structure of psychrotrophic strains. *Appl. Environ. Microbiol.* 72, 1569-1578.

Sorsa,L.J., Dufke,S., Heesemann,J., and Schubert,S. (2003). Characterization of an *iroBCDEN* gene cluster on a transmissible plasmid of uropathogenic *Escherichia coli*: evidence for horizontal transfer of a chromosomal virulence factor. *Infect. Immun.* 71, 3285-3293.

Sukhodolets,V.V. (2006). [Unequal crossing-over in *Escherichia coli*]. *Genetika* 42, 1526-1535.

Tenaillon,O., Skurnik,D., Picard,B., and Denamur,E. (2010). The population genetics of commensal *Escherichia coli*. *Nat. Rev. Microbiol.* 8, 207-217.

Thomson,N.R., Howard,S., Wren,B.W., Holden,M.T., Crossman,L., Challis,G.L., Churcher,C., Mungall,K., Brooks,K., Chillingworth,T., Feltwell,T., Abdellah,Z., Hauser,H., Jagels,K., Maddison,M., Moule,S., Sanders,M., Whitehead,S., Quail,M.A., Dougan,G., Parkhill,J., and Prentice,M.B. (2006). The complete genome sequence and comparative genome analysis of the high pathogenicity *Yersinia enterocolitica* strain 8081. *PLoS. Genet.* 2, e206.

Touchon,M., Hoede,C., Tenaillon,O., Barbe,V., Baeriswyl,S., Bidet,P., Bingen,E., Bonacorsi,S., Bouchier,C., Bouvet,O., Calteau,A., Chiapello,H., Clermont,O., Cruveiller,S., Danchin,A., Diard,M., Dossat,C., Karoui,M.E., Frapy,E., Garry,L., Ghigo,J.M., Gilles,A.M., Johnson,J., Le,B.C., Lescat,M., Mangenot,S., Martinez-Jehanne,V., Matic,I., Nassif,X., Oztas,S., Petit,M.A., Pichon,C., Rouy,Z., Ruf,C.S., Schneider,D., Tourret,J., Vacherie,B., Vallenet,D., Medigue,C., Rocha,E.P., and Denamur,E. (2009). Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. *PLoS. Genet.* 5, e1000344.

Toval,F., Kohler,C.D., Vogel,U., Wagenlehner,F., Mellmann,A., Fruth,A., Schmidt,M.A., Karch,H., Bielaszewska,M., and Dobrindt,U. (2014). Characterization of *Escherichia coli* isolates from hospital inpatients or outpatients with urinary tract infection. *J. Clin. Microbiol.* 52, 407-418.

Treangen,T.J. and Salzberg,S.L. (2012). Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat. Rev. Genet.* 13, 36-46.

Vimont,S., Boyd,A., Bleibtreu,A., Bens,M., Goujon,J.M., Garry,L., Clermont,O., Denamur,E., Arlet,G., and Vandewalle,A. (2012a). The CTX-M-15-producing *Escherichia coli* clone O25b: H4-ST131 has high intestine colonization and urinary tract infection abilities. *PLoS. One.* 7, e46547.

Vimont,S., Boyd,A., Bleibtreu,A., Bens,M., Goujon,J.M., Garry,L., Clermont,O., Denamur,E., Arlet,G., and Vandewalle,A. (2012b). The CTX-M-15-producing *Escherichia coli* clone O25b: H4-ST131 has high intestine colonization and urinary tract infection abilities. *PLoS. One.* 7, e46547.

Wiles,T.J., Kulesus,R.R., and Mulvey,M.A. (2008). Origins and virulence mechanisms of uropathogenic *Escherichia coli*. *Exp. Mol. Pathol.* *85*, 11-19.

Williams,K.P., Gillespie,J.J., Sobral,B.W., Nordberg,E.K., Snyder,E.E., Shallom,J.M., and Dickerman,A.W. (2010). Phylogeny of gammaproteobacteria. *J. Bacteriol.* *192*, 2305-2314.

Winterberg,K.M., Luecke,J., Bruegl,A.S., and Reznikoff,W.S. (2005). Phenotypic screening of *Escherichia coli* K-12 Tn5 insertion libraries, using whole-genome oligonucleotide microarrays. *Appl. Environ. Microbiol.* *71*, 451-459.

Wirth,T., Falush,D., Lan,R., Colles,F., Mensa,P., Wieler,L.H., Karch,H., Reeves,P.R., Maiden,M.C., Ochman,H., and Achtman,M. (2006). Sex and virulence in *Escherichia coli*: an evolutionary perspective. *Mol. Microbiol.* *60*, 1136-1151.

Zhou,J. and Rudd,K.E. (2013). EcoGene 3.0. *Nucleic Acids Res.* *41*, D613-D624.

Zinder,N.D. and Lederberg,J. (1952). Genetic exchange in *Salmonella*. *J. Bacteriol.* *64*, 679-699.

Danksagung

Die vorliegende Arbeit wurde im Zeitraum 03/2012 bis 06/2016 am Max von Pettenkofer-Institut der Ludwig-Maximilians-Universität München angefertigt.

Im Folgenden möchte ich allen herzlich danken, die einen Beitrag zur Anfertigung dieser Arbeit geleistet haben:

Zuerst möchte ich meinem Doktorvater Prof. Dr. Wolfgang Fischer und Prof. Dr. Sören Schubert für die Bereitstellung des Laborplatzes und der Materialien danken. Vor allem bedanke ich mich herzlich für die sehr gute Betreuung der vorliegenden Arbeit, die konstruktive Kritik und das Einbringen neuer Ideen.

Birgit Groß und Gabriele Liegl danke ich für die gute Arbeitsatmosphäre, für die Ordnung im Labor und für die Hilfestellungen bei den Versuchen.

Christina Hamacher und Maria Almagro-Moltó möchte ich für ein stets offenes Ohr und eine sehr gute Zeit im Labor danken.

Meinen Kollegen und guten Freunden Dr. Christian Meyer und Dr. Martin Arnold möchte ich herzlichst für die gute Zeit innerhalb und außerhalb des Labors danken. Vielen Dank Chris für die fachlichen und nicht-fachlichen Diskussionen, die Hilfe im Labor und alles drum herum!

Meinen Partnern und Kollegen des "ERA-NET PathoGenoMics" Konsortiums möchte ich für die guten Anregungen zu meinem Projekt und den schönen Aufenthalten in Paris, Münster und Wien danken.

Den Mitarbeitern der Diagnostik möchte ich für die gute Atmosphäre und die Mithilfe bei der materiellen Versorgung danken.

Vielen Dank auch an die netten Kollegen vom Max von Pettenkofer-Institut. Es war schön, die diversen Kicker-Turniere, Wiesn-Besuche und Feiern mit euch zu erleben.

Meinen guten Freunden Ulrike Mraß und Sebastian Schäfer möchte ich für die schönen Aktivitäten außerhalb der Arbeit danken. Ihr habt immer Zeit und stets ein offenes Ohr für mich!

Vielen herzlichen Dank an Verena Henn! Du hast mir immer Halt gegeben, wenn ich Dich gebraucht habe und mich immer unterstützt! Auch Deinen Eltern und Deiner Familie möchte ich an dieser Stelle für die schöne Zeit und die tollen Familienfeste danken!

Mein größter Dank geht an meine Familie, die mir immer geholfen und mir einiges gelehrt hat. Vielen Dank Cyril für die guten Abende! Besonders meiner Mutter danke ich für die guten Ratschläge und die unendliche Unterstützung!

Eidesstattliche Versicherung

Messerer, Maxim

Ich erkläre hiermit an Eides statt,

dass ich die vorliegende Dissertation mit dem Thema

"Untersuchungen des horizontalen Gentransfers in *Enterobacteriaceae* durch Next-Generation Sequencing"

selbstständig verfasst, mich außer der angegebenen keiner weiteren Hilfsmittel bedient und alle Erkenntnisse, die aus dem Schrifttum ganz oder annähernd übernommen sind, als solche kenntlich gemacht und nach ihrer Herkunft unter Bezeichnung der Fundstelle einzeln nachgewiesen habe.

Ich erkläre des Weiteren, dass die hier vorgelegte Dissertation nicht in gleicher oder in ähnlicher Form bei einer anderen Stelle zur Erlangung eines akademischen Grades eingereicht wurde.

München, den _____

Unterschrift Doktorandin/Doktorand